

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:09:11 ; Search time 192 Seconds
(without alignments)
3001.422 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773

Sequence: 1 MAGNSLVLPVLWGRKAPTH.....TSNRNVILMAHDGKEHRFMV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7773	100.0	1490	8	ADQ60212 Human rab
2	7772	100.0	1490	4	AAE10802 Human TGF
3	7763	99.9	1490	5	ABP70125 Human NOV
4	7418.5	95.4	1489	4	AAE10803 Mouse TGF
5	7373	94.9	1488	4	AAE10801 Rat TGF-b
6	3128.5	40.2	1525	4	ABBE6447 Drosophil
7	3128.5	40.2	1525	8	ADS96554 Drosophil
8	2805.5	36.1	559	8	ADA5388 Human pro
9	1442	18.6	775	7	ADM04809 Human pro
10	1003	12.9	209	4	ADM19863 Protein e
11	909	11.7	210	4	ADM19833 Protein e
12	332	4.3	81	4	ADM20106 Protein e
13	228.5	2.9	579	8	ADS29860 Bacterial
14	228.5	2.9	608	8	ADS41794 Bacterial
15	220.5	2.8	1005	8	ADS41812 Bacterial
16	205.5	2.6	626	8	ADS41817 Bacterial
17	202	2.6	1136	8	ADS30013 Bacterial
18	199	2.6	1240	4	ABBE1596 Drosophil
19	197.5	2.5	429	8	ADS22315 Bacterial
20	197	2.5	580	8	ADR86389 Aspergill
21	194.5	2.5	514	6	ABO07190 Human p53
22	191.5	2.5	478	8	ADS29872 Bacterial
23	187.5	2.4	459	4	ADM19860 Protein e
24	187.5	2.4	514	4	AAB95225 Human pro
25	187.5	2.4	514	8	ADJ75513 Marker ge

Query Match 100.0%; Score 7773; DB 8; Length 1490;

26	186	2.4	573	3	AAB53475	Aab53475 Human col
27	185.5	2.4	679	8	ADS41629	Aad41629 Bacterial
28	185	2.4	972	4	AAM24010	Aam24010 Hamster E
29	185	2.4	972	4	AAM23988	Aam23988 Hamster E
30	182	2.3	943	6	ABR53214	ABr53214 Protein s
31	182	2.3	943	7	ADK63734	Adk63734 Disease t
32	181.5	2.3	514	8	ADJ76263	Adj76263 Marker ge
33	178.5	2.3	1278	8	ABM80447	Abm80447 Tumour-as
34	178.5	2.3	1278	8	ADR96784	Adr96784 Human SRE
35	178	2.3	610	8	ADS41755	Ads41755 Bacterial
36	176.5	2.3	886	4	AAB93996	Aab93996 Human pro
37	174.5	2.2	579	8	ADM23920	Adm23920 Bacterial
38	174.5	2.2	587	2	AAV03204	Aay03204 Amino aci
39	173.5	2.2	1278	8	ADS10766	Ads10766 Human the
40	173	2.2	1481	7	ADC31541	Adc31541 Human nov
41	171	2.2	395	5	ABP51424	Abp51424 Human MDD
42	171	2.2	981	5	ABP73564	Abp73564 Candida a
43	169.5	2.2	358	4	ABG65223	Abg65223 Drosophil
44	169.5	2.2	1140	8	ADS30020	Ads30020 Bacterial
45	167.5	2.2	485	4	AAB68284	Aab68284 Amino aci

ALIGNMENTS

RESULT 1

ADQ60212
ID ADQ60212 standard; protein; 1490 AA.

XX ADQ60212;

XX 07-OCT-2004 (first entry)

XX Human rabconnectin 3-binding protein.

XX human; rabconnectin 3; GDP/GTP exchange reaction promotion protein;
binding protein; calcium dependent exocytosis.

XX Homo sapiens.

XX JP2004201673-A.

XX 22-JUL-2004.

XX 13-AUG-2003; 2003JP-00207500.

XX 01-NOV-2002; 2002JP-00319521.

XX (EISA) EISAI CO LTD.

XX WPI; 2004-556370/54.

XX N-PSDB; ADQ60211.

PT Novel protein that couples with rabconnectin 3 and GDP/GTP exchange
reaction promotion protein, useful for elucidating chemical substance
that activates or inhibits calcium dependent exocytosis.

PS Claim 1; SEQ ID NO 2; 52pp; Japanese.

CC The invention comprises the amino acid and coding sequence of a protein
that couples with rabconnectin 3 and GDP/GTP exchange reaction promotion
protein. The DNA and protein sequences of the invention are useful for
analysing immune tissue for locating the protein of the invention or for
analysing the expression level of the protein of the invention. The DNA
and protein sequences of the invention are useful for elucidating
chemical substances that activate or inhibit calcium dependent
exocytosis. The present amino acid sequence represents the human
rabconnectin 3 and GDP/GTP exchange reaction promotion protein-binding
protein of the invention.

SQ Sequence 1490 AA;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATVTCGHDGQICLMDLSVELQINPRALL	60
DB	1	MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATVTCGHDGQICLMDLSVELQINPRALL	60
QY	61	FHTASITCLSKACASDDKQYIVSASSEGEMCLWDVSDGRCIIFTKACHTHTGIQFYQFS	120
DB	61	FHTASITCLSKACASDDKQYIVSASSEGEMCLWDVSDGRCIIFTKACHTHTGIQFYQFS	120
QY	121	VGNQREGRLCHGHYPBILVVDATSLEVLVSVKSPDWISSMSIIRSHRTQEDTVVAL	180
DB	121	VGNQREGRLCHGHYPBILVVDATSLEVLVSVKSPDWISSMSIIRSHRTQEDTVVAL	180
QY	181	SVTGILKVLVITSEISDMQDTEPIFEESKPIYCONCOSISFCAFTORSLLVVCSSKYRW	240
DB	181	SVTGILKVLVITSEISDMQDTEPIFEESKPIYCONCOSISFCAFTORSLLVVCSSKYRW	240
QY	241	FDAGDYSLLCSGPFSENGQTWTGGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV	300
DB	241	FDAGDYSLLCSGPFSENGQTWTGGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV	300
QY	301	GKAVENLIPPVQHILLDRKDKEILLICPPVTRPFYGCREYFHKLIIQDSSGRNLINWISD	360
DB	301	GKAVENLIPPVQHILLDRKDKEILLICPPVTRPFYGCREYFHKLIIQDSSGRNLINWISD	360
QY	361	TADQSGSEGLAMTTSISLQEAQDKLNCPCAGIIDQLSVIPNSNEPLKVTASVVIPAHR	420
DB	361	TADQSGSEGLAMTTSISLQEAQDKLNCPCAGIIDQLSVIPNSNEPLKVTASVVIPAHR	420
QY	421	LVCGREDSIVIVPATQTAIVQLLQGEHMLRRGWPHTRTLGRHNKVTCLLYPHQVSARY	480
DB	421	LVCGREDSIVIVPATQTAIVQLLQGEHMLRRGWPHTRTLGRHNKVTCLLYPHQVSARY	480
QY	481	DQRYLSGGVDVFSVIWIDIFSGEMKHIFCVHGGEITQLLVPPENC SARVOHCICSVASDH	540
DB	481	DQRYLSGGVDVFSVIWIDIFSGEMKHIFCVHGGEITQLLVPPENC SARVOHCICSVASDH	540
QY	541	SVGLLSLREKCCIMLASRHLFPQIVIKWRSPDDYL VVCGSDGSYVYQMDTGALDRVCVG	600
DB	541	SVGLLSLREKCCIMLASRHLFPQIVIKWRSPDDYL VVCGSDGSYVYQMDTGALDRVCVG	600
QY	601	ITAVEIILNACDEAVPAADVLSLHPAVNLQAMTRRSLAALKNMAHKLQTLATNLLASEA	660
DB	601	ITAVEIILNACDEAVPAADVLSLHPAVNLQAMTRRSLAALKNMAHKLQTLATNLLASEA	660
QY	661	SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFFDVEALIIQLITTEASRPNTALISPENL	720
DB	661	SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFFDVEALIIQLITTEASRPNTALISPENL	720
QY	721	QKAGSSDKGGSFLTGKRAAVLFOQVKTETKENIKEHLDDDEEDEDEIMQREESDPEY	780
DB	721	QKAGSSDKGGSFLTGKRAAVLFOQVKTETKENIKEHLDDDEEDEDEIMQREESDPEY	780
QY	781	RSSKSPLTILEYNLWMDTAKLFMSCLHANGLEVLDEVCLDRGLMKPHCTVTSFGLLSR	840
DB	781	RSSKSPLTILEYNLWMDTAKLFMSCLHANGLEVLDEVCLDRGLMKPHCTVTSFGLLSR	840
QY	841	GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBGVGKGTGVSVRAVTTQHLLSIIISLANT	900
DB	841	GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBGVGKGTGVSVRAVTTQHLLSIIISLANT	900
QY	901	LMSMNTATFTIGDHMKKGTPPPRPSTPDLISKARGSPPTSSNIYQGVAKQVAAPVVSARS	960
DB	901	LMSMNTATFTIGDHMKKGTPPPRPSTPDLISKARGSPPTSSNIYQGVAKQVAAPVVSARS	960
QY	961	ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLIGLDKFRPPLLEMLARRWQDR	1020
DB	961	ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLIGLDKFRPPLLEMLARRWQDR	1020
QY	1021	LEVREAAQALLLAELRR IEQAGRKEAIDAWAPYLPOYIDHVIISPGVTSEAAQTITTPADA	1080

Db	1021	LEVREAQAALLLAELRRIEQAGRKEALDAPVLPQYIDHVIISPGVTSEAAQTITTPADA	1081
Qy	1081	SGPEAKVQEEHDLVDDDIITGCLSSVPQMKKISTSYEERKQATAIVLLGVIGABFGAE	1140
Db	1081	SGPEAKVQEEHDLVDDDIITGCLSSVPQMKKISTSYEERKQATAIVLLGVIGABFGAE	1140
Qy	1141	IETPKLLTRPRSSQIPEGGLTSGGSNYSLARHTCKALTFLQLQPPSPKLPPhstIRRT	1200
Db	1141	IETPKLLTRPRSSQIPEGGLTSGGSNYSLARHTCKALTFLQLQPPSPKLPPhstIRRT	1200
Qy	1201	AIDLIGRGFTVWEPYMDVSAVLMGLLELCADAEKQLANITMGPLSPAADSARSARHALS	1260
Db	1201	AIDLIGRGFTVWEPYMDVSAVLMGLLELCADAEKQLANITMGPLSPAADSARSARHALS	1260
Qy	1261	LIATARPPAFTITIAKEVHRHTALAAANTOSQONNHTTTTLARAKGEILRVIEILLERKPTD	1320
Db	1261	LIATARPPAFTITIAKEVHRHTALAAANTOSQONNHTTTTLARAKGEILRVIEILLERKPTD	1320
Qy	1321	VVDLLVEVMDIIMYCLEGSLVKKKGLOECPPAICRFYMSVSYERNHRIAVGARHGVSVALY	1380
Db	1321	VVDLLVEVMDIIMYCLEGSLVKKKGLOECPPAICRFYMSVSYERNHRIAVGARHGVSVALY	1380
Qy	1381	DIRTKQCTTHGHKGPITAVAFAPDGRYLATYNTDISHISFWQMTSLIGSIGMLNSAPO	1440
Db	1381	DIRTKQCTTHGHKGPITAVAFAPDGRYLATYNTDISHISFWQMTSLIGSIGMLNSAPO	1440
Qy	1441	LRCIKTQVPPVQSPASPGSHNALKLARLIWTSNRNVILMAHDCKEHRFMV	1490
Db	1441	LRCIKTQVPPVQSPASPGSHNALKLARLIWTSNRNVILMAHDCKEHRFMV	1490
RESULT 2			
AAE10802			
ID	AAE10802 standard; protein; 1490 AA.		
XX			
AC	AAE10802;		
XX			
DT	18-DEC-2001 (first entry)		
XX			
DE	Human TGF-beta Resistance Associated Gene (TRAG) protein.		
XX			
KW	Human; Transforming growth factor-beta; TGF-beta; cellular proliferation;		
KW	phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;		
KW	oncogenesis; gene mapping; transgenic animal; chromosome 18.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	9..47	
FT		/label= WD_repeat	
FT	Domain	56..95	
FT		/label= WD_repeat	
FT	Domain	239..246	
FT		/label= Tyrosine phosphorylation_motif	
FT	Domain	459..498	
FT		/label= WD_repeat	
FT	Domain	549..588	
FT		/label= WD_repeat	
FT	Domain	773..780	
FT		/label= Tyrosine phosphorylation_motif	
XX			
FN	WO200166739-A1.		
XX			
PD	13-SEP-2001.		
XX			
PF	12-FEB-2001; 2001WO-US0004475.		
XX			
PR	07-MAR-2000; 2000US-0187572P.		
PR	08-MAR-2000; 2000US-0187848P.		
XX			
FA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Thorgeirsson SS, Sanders S;		

XX WPI; 2001-589942/66.
DR N-PSDB; AAD18130.

XX Novel isolated polypeptide for studying and modulating mechanisms
PT involved in cellular proliferation comprises transforming growth factor-
PT beta resistance associated polypeptide fragment.

XX Claim 1; Page 56-61; 89pp; English.

XX The present sequence is human transforming growth factor (TGF)-beta
CC resistance associated (TRAG) protein. Human TRAG gene is located on
CC chromosome 18. TRAG DNA or protein is useful for studying and modulating
CC mechanisms involved in cellular proliferation, and for modulating
CC cellular phenotype. It is also used for screening aggressive metastasis
CC in cancer cells, and for targets used in cancer therapy. TRAG protein is
CC useful for evaluating factors that interact with and/or control TGF-beta
CC signalling for understanding both cell proliferation control and
CC oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful
CC as hybridisation probe, in chromosome and gene mapping, in the generation
CC of antisense RNA and DNA, for the preparation of TRAG protein by
CC recombinant techniques, and for generating either transgenic or knockout
CC animals which, in turn, are useful in the development and screening of
CC therapeutically useful reagents

XX Sequence 1490 AA;

Query Match 100.0%; Score 7772; DB 4; Length 1490;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1489; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDQIQCLWDSVELQINPRALL 60
DB 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDQIQCLWDSVELQINPRALL 60
QY 61 FGTASTTCLSKACASDQYIYSASESGEMCLWDVSDGRCIEFTKLACTHTTGIOYQFS 120
DB 61 FGTASTTCLSKACASDQYIYSASESGEMCLWDVSDGRCIEFTKLACTHTTGIOYQFS 120
QY 121 VGNQREGRLCHGHYPVLLVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
DB 121 VGNQREGRLCHGHYPVLLVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
QY 181 SVTGILKWTVTSEISDMQTEPIFEESKPIYQNCQSIQSPCAFTORSLLVVCYSKYRW 240
DB 181 SVTGILKWTVTSEISDMQTEPIFEESKPIYQNCQSIQSPCAFTORSLLVVCYSKYRW 240
QY 241 FDAGDYSLLCSGSPENGQTTGSDVSSDKVLIITWENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 FDAGDYSLLCSGSPENGQTTGSDVSSDKVLIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GKAVENLIPVQHILLDRKDKELLIIPPVTRFPVGCREYFPHKLLIOGDSGRLNIWNISD 360
DB 301 GKAVENLIPVQHILLDRKDKELLIIPPVTRFPVGCREYFPHKLLIOGDSGRLNIWNISD 360
QY 361 TADKQSGEGLAMTTSISLQAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGR 420
DB 361 TADKQSGEGLAMTTSISLQAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGR 420
QY 421 LVCGREGDSIVIVPATQTAIVQLLOQEHMLRRCWPHRTLRGRNKTCLLYPHQVSARY 480
DB 421 LVCGREGDSIVIVPATQTAIVQLLOQEHMLRRCWPHRTLRGRNKTCLLYPHQVSARY 480
QY 481 DQRYLSGGVDFSVIITWIDIFSGEMKHFVCHGGEITQLLPENCSARVQHCICSVASDH 540
DB 481 DQRYLSGGVDFSVIITWIDIFSGEMKHFVCHGGEITQLLPENCSARVQHCICSVASDH 540
QY 541 SVGLLSIREKKCITMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSGVYVQMDTGALDRCVMG 600
DB 541 SVGLLSIREKKCITMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSGVYVQMDTGALDRCVMG 600
QY 601 ITAVEILNACDEAVPAVDLSHPAVNLKQATRRSLAALKNWAHKLQTLATNLASEA 660
DB 601 ITAVEILNACDEAVPAVDLSHPAVNLKQATRRSLAALKNWAHKLQTLATNLASEA 660

DB 601 ITAVEILNACDEAVPAVDLSHPAVNLKQATRRSLAALKNWAHKLQTLATNLASEA 660
QY 661 SDKGNLPKYSHNSLWVQAIKTNLTDPDIHVLFPDVEALIIQLLTERASRNTALISPENL 720
DB 661 SDKGNLPKYSHNSLWVQAIKTNLTDPDIHVLFPDVEALIIQLLTERASRNTALISPENL 720
QY 721 QKASGSDKGGSLTGTGRAAVLFOQVKETIKENIKHLLDDEDEDEHMRREESDPEY 780
DB 721 QKASGSDKGGSLTGTGRAAVLFOQVKETIKENIKHLLDDEDEDEHMRREESDPEY 780
QY 781 RSSKSPKPLTLEVLNLTWDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVTSFGLLSR 840
DB 781 RSSKSPKPLTLEVLNLTWDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVTSFGLLSR 840
QY 841 GGHMSLMLPGYNOPACKLSHGKTEVGKLPASGEGVGKGTGVSVRAVTTQHLLSIISLANT 900
DB 841 GGHMSLMLPGYNOPACKLSHGKTEVGKLPASGEGVGKGTGVSVRAVTTQHLLSIISLANT 900
QY 901 LMSMTNATIGDHMKKGTPRPSPPTDLSKARGSPPTSSNIVOGQIKOVAAPVVSARSD 960
DB 901 LMSMTNATIGDHMKKGTPRPSPPTDLSKARGSPPTSSNIVOGQIKOVAAPVVSARSD 960
QY 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAWHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
DB 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAWHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
QY 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPLPQYIDHVISPGVTSEAAQTITAPDA 1080
DB 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPLPQYIDHVISPGVTSEAAQTITAPDA 1080
QY 1081 SGPEAKVQBEEDHLDVDDITTTGCLSSVPQMKKISTSEYERRRKOATAIIVLGVIGABFGAE 1140
DB 1081 SGPEAKVQBEEDHLDVDDITTTGCLSSVPQMKKISTSEYERRRKOATAIIVLGVIGABFGAE 1140
QY 1141 IEPPKLLTTPRSSSQIPEGFGLTSGGSNSYSLARHTCKALTFLLOPPSPKLPPHSTIRRT 1200
DB 1141 IEPPKLLTTPRSSSQIPEGFGLTSGGSNSYSLARHTCKALTFLLOPPSPKLPPHSTIRRT 1200
QY 1201 AIDLIGRGFTVMEPYMDVSAVLMGLLELCADEKQALANITMGLPSAADSARSARHALS 1260
DB 1201 AIDLIGRGFTVMEPYMDVSAVLMGLLELCADEKQALANITMGLPSAADSARSARHALS 1260
QY 1261 LIATARPAPFITTAKEVHRHTALAANTOSQONMHTTTLARAKEGLRVEILIEKPTD 1320
DB 1261 LIATARPAPFITTAKEVHRHTALAANTOSQONMHTTTLARAKEGLRVEILIEKPTD 1320
QY 1321 VDVLLEVMDDIIMYCLEGSLVKKKGLOECFPAICRFYMYVYERNHRIAVGARHGSVALY 1380
DB 1321 VDVLLEVMDDIIMYCLEGSLVKKKGLOECFPAICRFYMYVYERNHRIAVGARHGSVALY 1380
QY 1381 DIRTGKQTIHGHKGPITAVAPDGRYLATYSNTDHSISFWQMTSLGSGIGMLNSAPO 1440
DB 1381 DIRTGKQTIHGHKGPITAVAPDGRYLATYSNTDHSISFWQMTSLGSGIGMLNSAPO 1440
QY 1441 LRCIKTYQVPPVQSPGSHNALKARLIWTSNRNVILMAHDGKEHRFW 1490
DB 1441 LRCIKTYQVPPVQSPGSHNALKARLIWTSNRNVILMAHDGKEHRFW 1490

RESULT 3

ABP70125

ID ABP70125 standard; protein; 1490 AA.

XX

AC ABP70125;

XX

DT 27-JAN-2003 (first entry)

XX Human NOV36a.

DE

XX Human; anti-HIV; cytostatic; antidiabetic; antiaesthetic; cachexia; AIDS;

XX antinflammatory; cardiant; haemostatic; neuroprotective; anorectic;

XX nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;

XX antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;

KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
 KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
 KW immune disorder; haematopoietic disorder; cardiovascular disorder;
 KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
 KW metabolic syndrome X; wasting disorder; cell differentiation;
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis.
 XX

OS Homo sapiens.

XX WO200272771-A2.

XX 19-SEP-2002.

XX 08-MAR-2002; 2002WO-US007288.

XX 08-MAR-2001; 2001US-0274101P.

XX 08-MAR-2001; 2001US-0274194P.

XX 08-MAR-2001; 2001US-0274281P.

XX 08-MAR-2001; 2001US-0274322P.

XX 09-MAR-2001; 2001US-0274849P.

XX 12-MAR-2001; 2001US-0275235P.

XX 13-MAR-2001; 2001US-0275578P.

XX 13-MAR-2001; 2001US-0275579P.

XX 14-MAR-2001; 2001US-0275601P.

XX 14-MAR-2001; 2001US-0276000P.

XX 16-MAR-2001; 2001US-0276776P.

XX 19-MAR-2001; 2001US-0276994P.

XX 20-MAR-2001; 2001US-0277239P.

XX 20-MAR-2001; 2001US-0277321P.

XX 20-MAR-2001; 2001US-0277327P.

XX 20-MAR-2001; 2001US-0277338P.

XX 21-MAR-2001; 2001US-0277791P.

XX 22-MAR-2001; 2001US-0277833P.

XX 23-MAR-2001; 2001US-0278152P.

XX 26-MAR-2001; 2001US-0278894P.

XX 27-MAR-2001; 2001US-0278999P.

XX 27-MAR-2001; 2001US-0279036P.

XX 28-MAR-2001; 2001US-0279344P.

XX 30-MAR-2001; 2001US-0279995P.

XX 30-MAR-2001; 2001US-0280233P.

XX 02-APR-2001; 2001US-0280802P.

XX 02-APR-2001; 2001US-0280822P.

XX 02-APR-2001; 2001US-0280900P.

XX 04-APR-2001; 2001US-0281194P.

XX 13-APR-2001; 2001US-0283675P.

XX 30-APR-2001; 2001US-0287424P.

XX 02-MAY-2001; 2001US-0288056P.

XX 03-MAY-2001; 2001US-0288342P.

PR 03-DEC-2001; 2001US-0337426P.

PR 03-DEC-2001; 2001US-0338092P.

PR 04-DEC-2001; 2001US-0337185P.

PR 03-JAN-2002; 2002US-0345705P.

XX 08-MAR-2002; 2002US-00093463.

XX (CURA-) CURAGEN CORP.

XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;

PI Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM;

PI Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;

PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;

PI Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;

PI Zhong M;

XX WPI; 2002-732824/79.

DR N-PSDB; ABV99403.

XX New NOVX polypeptides and polynucleotides, useful for preventing,

PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,

PT Alzheimer's disease, dyslipidemia, obesity, immune or hematopoietic

PT disorders, and asthma.

XX Claim 1; Page 232; 619pp; English.

XX The present invention relates to new isolated proteins (NOVX) and their

CC coding sequences (ABV99327-ABV99535 and ABP70049-ABP70149), where X is

CC any number from 1 to 48. The NOVX proteins and coding sequences are

CC useful in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, preferably a NOVX-associated disorder.

CC The NOVX coding sequences and proteins are useful for treating,

CC preventing or diagnosing diseases such as metabolic disorders, diabetes,

CC obesity, infectious disease, anorexia, cancer-associated cachexia,

CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's

CC disease, immune disorders, haematopoietic disorders, cardiovascular

CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic

CC disturbances associated with obesity, metabolic syndrome X or wasting

CC disorders associated with chronic diseases or various cancers. The NOVX

CC coding sequences and proteins may also be used as targets for the

CC identification of small molecules that modulate or inhibit e.g.

CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,

CC wound healing and angiogenesis, in gene therapy, in generation of

CC antibodies that bind immunospecifically to NOVX substances for use in

CC therapeutic or diagnostic methods

XX Sequence 1490 AA;

SQ

Query Match 99.9%; Score 7763; DB 5; Length 1490;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1488; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAGNSLVLPVLWGRKAPTHCISAVLLTDDGATVTCGHDGQICLWDLVELQINPRALL 60
 DB 1 MAGNSLVLPVLWGRKAPTHCISAVLLTDDGATVTCGHDGQICLWDLVELQINPRALL 60
 QY 61 FGHASITCLSKACASSDKQIYVSASEGEMCLWVSDGRCIEFTKLACTHTGQIQYQFS 120
 DB 61 FGHASITCLSKACASSDKQIYVSASEGEMCLWVSDGRCIEFTKLACTHTGQIQYQFS 120
 QY 121 VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
 DB 121 VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
 QY 181 SVTGILKVMIVTSEISDMQDTEPIFEESKPIYQNCQCSISCAFTQRIQLLVCSKYWRV 240
 DB 181 SVTGILKVMIVTSEISDMQDTEPIFEESKPIYQNCQCSISCAFTQRIQLLVCSKYWRV 240
 QY 241 FDAGDYSLLCSGSENGQWTGGDFVSSDKVIWTENGOSYIYKLPASCLPASDSFRSDV 300
 DB 241 FDAGDYSLLCSGSENGQWTGGDFVSSDKVIWTENGOSYIYKLPASCLPASDSFRSDV 300
 QY 301 GKAVENLIPPVQHILLDRKDKELLICPPVTRFPYFGREYFHKLLIOGDSGRLNIWNISD 360

Db 301 GKAVENLIPVQHILLDRKDKELLI CPVTRFFVGCBEYFHKLLIQDSSGRNLWNISD 360
Qy 361 TADKQSEGLAMTTSISLQBAFDKLPNCPAGIITDQLSVIPNSNEPLKVTASVYPAHGR 420
Db 361 TADKQSEGLAMTTSISLQBAFDKLPNCPAGIITDQLSVIPNSNEPLKVTASVYPAHGR 420
Qy 421 LVCGRDGSIVIVPATOTAI VOLLQGEHMLRGHPHRTLRGHRNKVTCLLYPHQVSARY 480
Db 421 LVCGRDGSIVIVPATOTAI VOLLQGEHMLRGHPHRTLRGHRNKVTCLLYPHQVSARY 480
Qy 481 DQRLISGGVDFSVI IWDIFSGEMKHIFCVHGGBITOLLVPPENC SARVOHCICSVASDH 540
Db 481 DQRLISGGVDFSVI IWDIFSGEMKHIFCVHGGBITOLLVPPENC SARVOHCICSVASDH 540
Qy 541 SVGLLSLREKCKIMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSVYVWQMDTGALDRCVMG 600
Db 541 SVGLLSLREKCKIMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSVYVWQMDTGALDRCVMG 600
Qy 601 ITAVEILLNACDEAVPAAVDSLSHPAVNLKQAMTRRS LAALKNMAHHKLOTATNLLASEA 660
Db 601 ITAVEILLNACDEAVPAAVDSLSHPAVNLKQAMTRRS LAALKNMAHHKLOTATNLLASEA 660
Qy 661 SDKGNLPKYSHNSLWQAIKTNITDPDIHVLFPDVEALIIQLLTERASRPNTALISPENL 720
Db 661 SDKGNLPKYSHNSLWQAIKTNITDPDIHVLFPDVEALIIQLLTERASRPNTALISPENL 720
Qy 721 QKASGSSDKGSGFLTGKRAAVLFOQVKETIKENI KEHLDDDEEDEBEIMQRRESDPEY 780
Db 721 QKASGSSDKGSGFLTGKRAAVLFOQVKETIKENI KEHLDDDEEDEBEIMQRRESDPEY 780
Qy 781 RSSKSKPLTLEVLNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVTSFGLLSR 840
Db 781 RSSKSKPLTLEVLNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVTSFGLLSR 840
Qy 841 GGHMSLMLPGVNOPACKLSCKTEVGRKLPASEGVGKGTGVSRVAVTTOHLLSTIISLANT 900
Db 841 GGHMSLMLPGVNOPACKLSCKTEVGRKLPASEGVGKGTGVSRVAVTTOHLLSTIISLANT 900
Qy 901 LMSMTNATFGDHMKKGTPRPSPDLSKARGSPPTSSNI VQGQIKQVAAAPVVSARS 960
Db 901 LMSMTNATFGDHMKKGTPRPSPDLSKARGSPPTSSNI VQGQIKQVAAAPVVSARS 960
Qy 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDRC 1020
Db 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDRC 1020
Qy 1021 LEVREAAQALLAELEIEQGRKEADAWAPYLPOVIDHVI SPGVTSBAAOITITAPDA 1080
Db 1021 LEVREAAQALLAELEIEQGRKEADAWAPYLPOVIDHVI SPGVTSBAAOITITAPDA 1080
Qy 1081 SGPEAKVQEBEHLVDVDDITTCGLSSVPQMKKISTSYEERRKQATAILLVIGVIGAEFGAE 1140
Db 1081 SGPEAKVQEBEHLVDVDDITTCGLSSVPQMKKISTSYEERRKQATAILLVIGVIGAEFGAE 1140
Qy 1141 IEPKLLTRPRSSQIPEGFLTSGGNSYLARHTCKALFTLLQPPSPKLPHPHSTIRRT 1200
Db 1141 IEPKLLTRPRSSQIPEGFLTSGGNSYLARHTCKALFTLLQPPSPKLPHPHSTIRRT 1200
Qy 1201 AIDLIGGFTWPEYMDVSAVLMGLLELCADEKQALANI TGLPLSPAADSARSARHLS 1260
Db 1201 AIDLIGGFTWPEYMDVSAVLMGLLELCADEKQALANI TGLPLSPAADSARSARHLS 1260
Qy 1261 LIATARPPAFTITIAKEVHRHTALAANTQSQNNHTTTLARAKEILRVIEILIEKMPD 1320
Db 1261 LIATARPPAFTITIAKEVHRHTALAANTQSQNNHTTTLARAKEILRVIEILIEKMPD 1320
Qy 1321 VDLLEWMDIMYCLEGSLVKKKGLQECFPAICRFYVWSYERNHRIAVGARHGSVALY 1380
Db 1321 VDLLEWMDIMYCLEGSLVKKKGLQECFPAICRFYVWSYERNHRIAVGARHGSVALY 1380
Qy 1381 DIRTKCQTHGHKGPITAVAPADGRYLATYNTDSHISFWQNTSLGSGMLNSAPQ 1440
Db 1381 DIRTKCQTHGHKGPITAVAPADGRYLATYNTDSHISFWQNTSLGSGMLNSAPQ 1440

Qy 1441 LRCIKTYQVPPVQSPASGSHNALKARLIWTSNRNVILMAHDGKEHRFMV 1490
Db 1441 LRCIKTYQVPPVQSPASGSHNALKARLIWTSNRNVILMAHDGKEHRFMV 1490
RESULT 4
AAE10803
ID AAE10803 standard; protein; 1489 AA.
XX AAE10803;
XX 18-DEC-2001 (first entry)
XX Mouse TGF-beta Resistance Associated Gene (TRAG) protein.
XX Mouse; Transforming growth factor-beta; TGF-beta; cellular proliferation;
XX phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;
XX oncogenesis; gene mapping; transgenic animal; chromosome 18.
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Domain 9..47
XX FT /label= WD_repeat
XX Domain 56..95
XX FT /label= WD_repeat
XX Domain 239..246
XX FT /label= Tyrosine phosphorylation_motif
XX Domain 459..498
XX FT /label= WD Repeat
XX Domain 549..588
XX FT /label= WD_repeat
XX Domain 773..780
XX FT /label= Tyrosine phosphorylation_motif
XX
XX WO200166739-A1.
XX 13-SEP-2001.
XX 12-FEB-2001; 2001WO-US004475.
XX 07-MAR-2000; 2000US-0187572P.
XX 08-MAR-2000; 2000US-0187848P.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Thorgeirsson SS, Sanders S;
XX WPI; 2001-589942/66.
XX N-PSDB; AAD18131.
XX Novel isolated polypeptide for studying and modulating mechanisms
XX involved in cellular proliferation comprises transforming growth factor-
XX beta resistance associated polypeptide fragment.
XX Claim 1; Page 62-66; 89pp; English.
XX The present sequence is mouse transforming growth factor (TGF)-beta
XX resistance associated (TRAG) protein. Mouse TRAG gene is located on
XX chromosome 18. TRAG DNA or protein is useful for studying and modulating
XX mechanisms involved in cellular proliferation, and for modulating
XX cellular phenotype. It is also used for screening aggressive metastasis
XX in cancer cells, and for targets used in cancer therapy. TRAG protein is
XX useful for evaluating factors that interact with and/or control TGF-beta
XX signalling for understanding both cell proliferation control and
XX oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful
XX as hybridisation probe, in chromosome and gene mapping, in the generation
XX of antisense RNA and DNA, for the preparation of TRAG protein by
XX recombinant techniques, and for generating either transgenic or knockout
XX animals which, in turn, are useful in the development and screening of
XX therapeutically useful reagents
XX

SQ	Sequence 1489 AA;	
Query Match	95.4%; Score 7418.5; DB 4; Length 1489;	
Best Local Similarity	95.0%; Pred. No. 0;	
Matches 1415; Conservative	36; Mismatches 38; Indels 1; Gaps 1;	
QY	1 MAGNSVLPTLVLMGRKAPTHCISAVLLTDDGATVITGCHDGOICLDVLSVELQINRALL 60	
DB	1 MAGNSVLPTLVLMGRKAPTHCISAILLTDGGTIVTGDHGOICLDVLSVELVFNPRALL 60	
QY	61 FGTASITCLSKACASDKQYIYVASSESGMCLWDVSDGRCIBFTKLACTHTGTGQYQFS 120	
DB	61 FGTASITCLSKACASDKRYTYVASANGEMCLWDVNDGRCIBFTKLACTHTGTGQYQFS 120	
QY	121 VGNQREGRLCHGHYPEILVVDATSLVLSYLSKISPDWISSMSIIRSHRTQEDTVVAL 180	
DB	121 VGNQREGRLCHGHYPEILVVDATSLVLSYLSKISPDWISSMSIIRSHRTQEDTVVAL 180	
QY	181 SVTGILKWTIVTSEISDMQDTEPIFEESKPIYQNCQSIQSPCAFTORSLLVVCYSKYWRV 240	
DB	181 SVTGILKWTIVTSENGMQDTEPIFEESKPIYQNCQSIQSPCAFTORSLLVVCYSKYWRV 240	
QY	241 FDAGDYSLLCSGFSNGQWTGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300	
DB	241 FDAGDYSLLCSGFSNGQWTGDFVSADKVIWTENGQSYIYKLPASCLPASDSFRSDV 300	
QY	301 GKAVENLIIPVQHLLDRKDKELLICPPVTRFPYGCREYFHKLLIQDSSGRLNIWNISD 360	
DB	301 GKAVENLIIPVQSHLLDQDKELVICPPVTRFPYGCKEYLHKLLIQDSSGRLNIWNISD 360	
QY	361 TADQSGSEGLAMTTSISLQEARFDKLPKPCAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420	
DB	361 IAEQBADEGLKMTTCLISQEARFDKLPKPCAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420	
QY	421 LVCGREDGSIIVPATQTAIVQLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY 480	
DB	421 LVCGREDGSIIVPATQTAIVQLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY 480	
QY	481 DQRYLSGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHICISVASDH 540	
DB	481 DQRYLSGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHICISVASDH 540	
QY	541 SVGLLSIREKKCTMLASRLHFPQVIKWRPSDDYLVVCGSDGVVYVQMDTGALDCVMG 600	
DB	541 SVGLLSIREKKCTMLASRLHFPQVIKWRPSDDYLVVCGSDGVVYVQMDTGALDCVMG 600	
QY	601 ITAVEILNACDEAVPAVDLSHPAVNLKQAMTFRSLAALKNNMAHKLQTLATNLLASEA 660	
DB	601 ITAVEILNACDEAVPAVDLSHPAVNLKQAMTFRSLAALKNNMAHKLQTLATNLLASEA 660	
QY	661 SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTEASRPNTALISPENL 720	
DB	661 SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTEASRPNTALISPENL 720	
QY	721 QKASGSDKGSFLTGTGRAAVLFQOVKETIKENIKHLLDDEEIMRQREESDPEY 780	
DB	721 QKASGSDKGSFLTGTGRAAVLFQOVKETIKENIKHLLDDEEIMRQREESDPEY 780	
QY	781 RSSKSKPLTLLEYNLMTDTAKLFMSCLHAWGLNEVLDEVCLDLGLMKPHCTVTSFGLLR 840	
DB	781 RASKSKPLTLLEYNLMTDTAKLFMSCLHAWGLNEVLDEVCLDLGLMKPHCTVTSFGLLR 840	
QY	841 GGHMSLMLPGYNQPAKLSHGKTEVGRKLPASGCVGKGTGVGRVATTQHLLSIISLANT 900	
DB	841 GGHMSLMLPGYNQAAGKLLHAKAEVGRKLPAAEGVGKGTGVGRVATTQHLLSIISLANT 900	
QY	901 LMSMTNATFGDMMKGPTRPPPTDLSKARGSPPTSSNIVQGOIKOVAAPVWSARSD 960	
DB	901 LMSMTNATFGDMMKGPTRPPPTDLSKARDSPPASSNIVQGOIKQAAAAPVWSARSD 960	
QY	961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAMHCVMLPDLLGLDKFRPPLLEMLARRWQDR 1020	
DB	961 ADHSGSDSAS-PALPTCFLVNEGWSQLAAMHCVMLPDLLGLERFRPPLLEMLARRWQDR 1019	

QY	1021 LEVREAAQALLAEELRRIEQAGRKEAIDAWAPVLPQYIDHVISPGVTSEAAOTITTAPDA 1080	
DB	1020 LEVREAAQALLAEELRRIEQAGRKETIDTWAYLPYQYMDHVISPGVTAGAMQTAAPDA 1079	
QY	1081 SGPEAKVQSEEHDLVDDDDITTCGLSSVPQMKKISTSYEERRKQATAI VLLGVIGAEFGAE 1140	
DB	1080 SGPEAKVQSEEHDLVDDDDITAGCLSSVPQMKKISTSYEERRKQATAI VLLGVIGAEFGAE 1139	
QY	1141 IEPPKLLTRPRSSQIPGFGLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPHSTIRT 1200	
DB	1140 IEPPKLLTRPRSSQIPGFGLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPHSTIRT 1199	
QY	1201 ATDLTGRGTVPWEPMYDVSALVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1260	
DB	1200 ATDLTGRGTVPWEPMYDVSALVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1259	
QY	1261 LIATARPAPAFITTIAREVHRHTALAANTOSQOMHTTTTLARAKGEILRVIELIEKMPD 1320	
DB	1260 LIATARPAPAFITTIAREVHRHTALAANTOSQOSIHTTTTLARAKGEILRVIELIEKMPD 1319	
QY	1321 VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRFYMYVSYTERNHRITAVGARHGSVALY 1380	
DB	1320 VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRFYMYVSYTERSHRITAVGARHGSVALY 1379	
QY	1381 DIRTGKCTQIHGKGPITAVAFAPDGRVLYATYSNTDSDHISFWQMTSLLSIGMLNSAPQ 1440	
DB	1380 DIRTGKCTQIHGKGPITAVAFAPDGRVLYATYSNTDSDHISFWQMTSLLSIGMLNSAPQ 1439	
QY	1441 LRCIKTYQVPPVQSPASPSGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV 1490	
DB	1440 LACIKTYQVPPVQSPASPSGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV 1489	
RESULT 5		
AAE10801		
ID	AAE10801 standard; protein; 1488 AA.	
AC	AAE10801;	
XX		
DT	18-DEC-2001 (first entry)	
XX		
DE	Rat TGF-beta Resistance Associated Gene (TRAG) protein.	
XX		
KW	Rat; Transforming growth factor-beta; TGF-beta; cellular proliferation; phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;	
KW	oncogenesis; gene mapping; transgenic animal.	
XX		
OS	Rattus sp.	
XX		
FT	Key	Location/Qualifiers
FT	Domain	9..47
FT		/label= WD_repeat
FT	Domain	56..95
FT		/label= WD_repeat
FT	Domain	239..246
FT		/label= Tyrosine_phosphorylation_motif
FT	Domain	459..498
FT		/label= WD_repeat
FT	Domain	549..588
FT		/label= WD_repeat
FT	Domain	773..780
FT		/label= Tyrosine_phosphorylation_motif
XX		
PN	WO200166739-A1.	
XX		
PD	13-SEP-2001.	
XX		
PF	12-FEB-2001; 2001WO-US004475.	
XX		
PR	07-MAR-2000; 2000US-0187572P.	
PR	08-MAR-2000; 2000US-0187848P.	
XX		

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Thorgeirsson SS, Sanders S;

WPI; 2001-589942/66.

N-PSDB; AAD18129.

Novel isolated polypeptide for studying and modulating mechanisms involved in cellular proliferation comprises transforming growth factor-beta resistance associated polypeptide fragment.

Claim 1; Page 51-56; 89pp; English.

The present sequence is rat transforming growth factor (TGF)-beta resistance associated (TRAG) protein. TRAG DNA or protein is useful for studying and modulating mechanisms involved in cellular proliferation, and for modulating cellular phenotype. It is also used for screening aggressive metastasis in cancer cells, and for targets used in cancer therapy. TRAG protein is useful for evaluating factors that interact with and/or control TGF-beta signalling for understanding both cell proliferation control and oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful as hybridisation probe, in chromosome and gene mapping, in the generation of antisense RNA and DNA, for the preparation of TRAG protein by recombinant techniques, and for generating either transgenic or knockout animals which, in turn, are useful in the development and screening of therapeutically useful reagents

XX SQ Sequence 1488 AA;

Query Match 94.9%; Score 7373; DB 4; Length 1488;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 1408; Conservative 39; Mismatches 41; Indels 2; Gaps 2;

QY 1 MAGNSLVLPVLWGRKAPTHCISAVLITDDGATIVTGDGQICLWDLVELQINPRALL 60
 DB 1 MAGNSLVLPVLWGRKAPTHCISILLTDDGATIVTGDGQICLWDLSEELVNPRA 60
 QY 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGIGYQFS 120
 DB 61 FGHATAITCLSKACASDQYIVSASANGEMCLWVSDGRCIEFTKLACTHTGIGYQFS 120
 QY 121 VGNQREGLLCHHYPEILVVDATSLVLSYLSKISPMWISSMSIIRSHRTQEDTVVAL 180
 DB 121 VGNQREGLLCHHYPEILVVDATSLVLSYLSKISPMWISSMSIIRSHRTQEDTVVAL 180
 QY 181 SVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQSFCAFTQSLVVCYSKYRW 240
 DB 181 SVTGILKWIIVTSEISGLQDTEPIFEESKPIYQNCQSFCAFTQSLVVCYSKYRW 240
 QY 241 FDAGDYSLLCSGSPENGQWTGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300
 DB 241 FDAGDYSLLCSGSPEDGQWTGDFVSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300
 QY 301 GKAVENLIIPVQHILDRKDKELLICPPVTRFPVGCREFPHKLLIQDSSGRININISD 360
 DB 301 GKAVENLIIPVQSHLLDKQDRELVICPPVTRFPVGCREFPHKLLIQDSSGRISINISD 360
 QY 361 TADKQSGEGLMTWTGISLOEAFDKNPCPAGIIDQLSVTPNSNEPLKVTASVIIPAKGR 420
 DB 361 IADKQNEGLKTTTCISIQDAFDKIKPCPAGIIDQLSVTPNSNEPLKVTASVIIPAKGR 420
 QY 421 LVCGREDGSIIVIPATQTAIVQLQGEHMLRRCWPPHRTLGRHNKVTCLLYPHQVSARY 480
 DB 421 LVCGREDGSIIVIPATQTAIVQLQGEHMLRRCWPPHRTLGRHNKVTCLLYPHQVSARY 480
 QY 481 DQRYLLSGGVDFSVIIVDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHCIVASDH 540
 DB 481 DQRYLLSGGVDFSVIIVDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHCIVASDH 540
 QY 541 SVGLLSLRKCKIMLASRHLFPQVTKWRPSDDLVVVGCSGDSVYVQMDTGALDRCVNG 600
 DB 541 SVGLLSLRKCKIMLASRHLFPQVTKWRPSDDLVVVGCTDGSVCVQMDTGALDRCVNG 600

QY 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSIALAKNMAHKLQTLATNLASEA 660
 DB 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSIALAKNMAHKLQTLATNLASEA 660
 QY 661 SDKGNLPKYSNLSMWQALKTNLTDPIHVLFPDVEALIIQLLTERASRNTALISPENL 720
 DB 661 SDKGNLPKYSNLSMWQALKTNLTDPIHVLFPDVEALIIQLLTERASRNTALISPENL 720
 QY 721 QKASGSSDKGGSFLLTGKRAAVLFOQVKETIKENIKHLLDDEDEBEIMQRREESDPEY 780
 DB 721 QKASGSSDKGGSFLLTGKRAAVLFOQVKETIKENIKHLLDDEDEBEIMQRREESDPEY 780
 QY 781 RSKSKPLLELLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSPGLLSR 840
 DB 781 RSKSKPLLELLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSPGLLSR 840
 QY 841 GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASBGVCKGTGVSVRAVTTQHLLSIIISLANT 900
 DB 841 GGHMSLMLPGYNOPACKLQAKAERKGPATESVGKGTITVSVRAVTTQHLLSIIISLANT 900
 QY 901 LMSMTNATFIGDHMKKGTPRPPRPTDLSKARGSPPTSSNIVQGOIKQVAAFPVVSARSD 960
 DB 901 LMSMTNATFIGDHMKKGTPRPPRPTDLSKARDSPPASSNIVQGOIKQAAAAP-VSARSA 959
 QY 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDRC 1020
 DB 960 ADHSGS-ASASPALRTCFLVNEGWSQLAAHCVMLPDLLGLGKFRPPLLEMLARRWQDRC 1018
 QY 1021 LEVREAAQALLAELEIRIEQAGRKRAIDAWAPVLPQYIDHVTSPGVTSEAAQTITAPDA 1080
 DB 1019 LEVREAAQALLAELEIRIEQAGRKETIDTWPVLPQYMDHVISPGBTAEAMQTAAPDA 1078
 QY 1081 SGPEAKVQBEEDHLDVDDDDITTCGLSSVPQMKKISTSYEBRRRKOATAIVLLGVIGABFGAE 1140
 DB 1079 SGPEAKVQBEEDHLDVDDDDITTCGLSSVPQMKKISTSYEBRRRKOATAIVLLGVIGABFGAE 1138
 QY 1141 IBPPKLLTPRPSQIPEGFGLTSGSNYSLARHTCKALTFLLLQPPSPKLPHPHSTIRRT 1200
 DB 1139 IBPPKLLTPRPSQIPEGFGLTSGSNYSLARHTCKALTFLLLQPPSPKLPHPHSTIRRT 1198
 QY 1201 AIDLIGRGFTVWEPMYDVSALVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1260
 DB 1199 AIDLIGRGFTVWEPMYDVSALVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1258
 QY 1261 LIATARPAPAFITIAKEVHRHTALAANTOSQONMHTTTTLARAKEILRVIEILIEKMPDT 1320
 DB 1259 LIATARPAPAFITIAKEVHRHTALAANTOSQONMHTTTTLARAKEILRVIEILIEKMPDT 1318
 QY 1321 VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRPFYVSYERNHRIAVGARHGVSVALY 1380
 DB 1319 VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRPFYVSYERNHRIAVGARHGVSVALY 1378
 QY 1381 DIRTGCKQTIHGHKGPIITAVAPDGRYLATVNTDTSHTSFQWMTSLGSGTGLNSAPQ 1440
 DB 1379 DIRTGCKQTIHGHKGPIITAVAPDGRYLATVNTDTSHTSFQWMTSLGSGTGLNSAPQ 1438
 QY 1441 LFCIKTYQVPPVQPAQSPGSHNALKLARLITSNRNVLMAHDKGKEHFRV 1490
 DB 1439 LFCIKTYQVPPVQPAQSPGSHNALKLARLITSNRNVLMAHDKGKEHFRV 1488

RESULT 6

ABB66447

ID ABB66447 standard; protein; 1525 AA.

XX ABB66447;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 26133.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.
XX PN W0200171042-A2.
XX PD
XX PF 27-SEP-2001.
XX XX 23-MAR-2001; 2001WO-US009231.
XX XX 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX XX (PEKE) PE CORP NY.
XX PA
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR N-PSDB; ABL10550.
XX DR
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 26133; 21pp + Sequence Listing; English.
XX PS
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signaling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1525 AA;

Query Match 40.2%; Score 3128.5; DB 4; Length 1525;
Best Local Similarity 41.0%; Pred. No. 6e-278;
Matches 655; Conservative 303; Mismatches 458; Indels 183; Gaps 32;

QY 1 MAGNSVLPLVTLVGRKAPHCISAVLLTDDGATVITGCHDQQLCLWDLS-VELQINPRAL 59
DB 1 MVTNLVVPVLMGPTAPHCISVFLSDDDQFLVITGCVYDQQLCLQVPEITLKMFRCL 60
QY 60 LFGHTASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLACTHTGIIQYQF 119
DB 61 LVGHSAPVLCVRLPENNPLVSSSENGEMCTWDLTDGKMEAVKLPQVHTQIQSYH- 119
QY 120 SVGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVA 179
DB 120 -TANSEDLVRLFCIGYAEIMVDPFSLVLYVLSKYPDWISAIHVLPRMRKDDVLA 178
QY 180 LSVTGLKWIIVTSETSDMDTEPIEERSKPIYCONCOSISFCAPTORSLLVVCSKYWR 239
DB 179 ITTGTGVKWLITG--NENHAPPIYENESKEIRCLNAITMNCQAQNRVLVLCYKIQW 236
QY 240 VFDAGDYSLVCSGSENGQTWTGDFVSSDKVLIITWENGOSYIYKLPASCLPASDSFRSD 299
DB 237 IYDAGDFTVLCVSIAPARERWQGGDFITSDRVMLWTDGKGLYKLPANCI PDKNEFHS- 295
QY 300 VGKAVENLIPVQHILLDRDKELLICPPVTRFFYGCREFPHKLLTQDSSGRINW--- 356
DB 296 --KSVVRDAPLYLVYVLOHAGDK-VLSCPPAMKLLQAGGQHN--LLRGDSEGYISVMNVP 350
QY 357 -----NLSADPADKGSSEGLAMTTSLSQEAFLKLANPCPAGIIDOLSVIPNSNEPLKVT 411
DB 351 EVELDNLISIILOAKMPRPKPHVCTSLVSEANIMDPPVPIIDQLSRITES--PVKLTS 408
QY 412 SVYIPAGRLVCGREDGSGIVIVPATQTAIVOLLQGEHMLRRGPPHRTLGRHKYTCLL 471
DB 409 SIYLPQOSRLVIGREDGSGIVIVPATQVVMQLLVGVIKQNFSDWFSHQILYGHGRVNCLL 468

QY 472 YPHQVSARYDORYLISGGVDRPSVIWIDIFSGBMKHFVCHGGEITQLLVPPECNCAVQH 531
DB 469 CFSMTHSRYEKSHLSGGIDFAVCLWDLYSGSLLHRFCVHAGEITQLLVPPECNCAVQH 528
QY 532 CICSVASHSVGLSLREKKCICMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSGVYVQMMDT 591
DB 529 CICSVASHSVTLVSLQERKCVTLASRHLFPVVTIKWRPLDDFLVGCSDGSGVYVQMMDT 588
QY 592 GALDRCVMGITAVEILNACDE-----AVPAADVLSLHAPVNLQAMTRSL 637
DB 589 GHLDRLVHGLMAEVLSCDEQAEDGGGGGSGNGASASEMGMANPAVHFPRGKSRNM 648
QY 638 AALKMAHHKLOTLATNLLASASDKGN--LPKYSHNSLMVQALTKTNLTDPDIHVLFPD 694
DB 649 NAIR-----HATORGITOLQOQHNGQNFDFLMKHSNPLVIGLRTNPKDAESHLFPD 704
QY 695 VEALIIQLITEASRPNTALISPENLOKASGSSDKGSFLTGKRAAVLFQVVK----- 747
DB 705 IESGLIFELHSEBYAQWTPATLESGLVHLQNPDKGSKMHLDAASKIGDPFNKVKNAVDVE 764
QY 748 ETIKENIKHLLDDEEEDIMR-----QRRESDPYRSKSKSPKPLTILLEYNLTM 797
DB 765 KILDKDKHGLVQKFKETEIYEKKVQAKVESLQKAVEPHEEQDLKSKIASKMEVTHM 824
QY 798 DTAKLFMSCLHAWGLNEVLDECLDRLGMLKPHCTVSPGLSLRGHMSLMLPGYNQ---- 853
DB 825 EYVALLSLHSLHSGDLPDLKMCETRLGLLRIVIVISYGLVSKAGYMSLLLTWQNNYAI 884
QY 854 -PACKL--SHGTEYGRKLPASEG-----GKGTYGVSRAVTTQHLHSIISLANTLMSMT 905
DB 885 PFGIQLPSSSKKRLPEELQRLHLEHTAVFTSLHWELSTLTNTNHLALVAMSNLILSMS 944
QY 906 NATPTGDMKMGTPRPRSTPDLSKARGSPPTSSNIVGOIKQVAAPVVSARSADHSG 965
DB 945 AASFL-----PDSEKHK-----KLQRLA-----QRTDSTLSEN 971
QY 966 SPSPSAPALHTCFLVNEGWSQLAAHCVMLPD---LLGLDKPRPPLLEMLAARWDCRLE 1022
DB 972 EERELMAHHSIQIKHWSLLATHCHFLLPDKIEALEPKPKRQVEMVKKWQHHCIE 1031
QY 1023 VREAAQALLLARRIEQAGRKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAP----- 1078
DB 1032 IREAAQILLGLTRMGKGRQLVESWAOYLPVLYTH--TEPIVGAQOQALALISQASGG 1089
QY 1079 -----DASGPEAKV-----QEEHDLVDDDDITTCCLSVSPQM 1110
DB 1090 AGSGSGNGGVGVGSGGGAGSGGPGGSGVPGDAHODEYEEEEEI-----I 1139
QY 1111 KKISTSYBERRKOATAI VLLGVIGAEFGAEI--EPPK-----LITRPRS 1152
DB 1140 RPSLSSELRKQTTAVILLGVIGAEFGQDI SQESPNHRGSI SMATGANLTSGVAGGER 1199
QY 1153 SSQIPGFGLTSGGNSYLARHTKALTFLLQPPSPKLPHPHSTIRRTAIDLIGRFTVM 1212
DB 1200 KSSVVEGFGIAN-----NLARLTSNALAHLLVAPPSPKLPQVTPLRRAAIDLIGRFTVM 1254
QY 1213 EPMVDSAVLMGLLELCADEAKQLANITMGLPLSPAADSARSARHLSLIATARPAPFIT 1272
DB 1255 EPLYDYSKVLGLLEISCEG-KAVPNLNTKPLTPQADACRTARHARLRIATARPAPFIT 1313
QY 1273 TTAKEVHRHTALAAANTQSQNMHT-TTLARAKEGLIRLVEILIEKMDPTDWDLLVEVMDI 1331
DB 1314 THAREVARVNTWQNAQSIINTPLTQSVLHAKAGEILQCVEMLDKMQSHIAGLLVEVMDI 1373
QY 1332 IMVLEGLSVKKGLQECPPAICRFYMVSYERNHRIAVGARHGSVALYDIFRGKQCTIH 1391
DB 1374 ALHCVDGNELKNRGLAELCPALCKFNQISHCATRIAVGANGSNLAIVELRQNKQCMIP 1433
QY 1392 GHKGPITAVAPADGRLYATYNTDISHISFWQNTSLLGSI GMLNSAPQLRCIKTYQVPP 1451
DB 1434 AHTHPITSLAFSPDGKGLVLSYCAENRLSPWOTSTGMFG-LGQ-----SOTRCTKGYSTAP 1488

QY 1452 VQASPGSHVAKLARIWTSNRVILMAHDKGKEHFMV 1490
Db 1489 IPDVS--RLNPMRLAKLVWNNRTVTLMLADGSETRENV 1525

RESULT 7

ADSS96554
ID ADS96554 standard; protein; 1525 AA.

XX AC ADS96554;

XX DT 02-DEC-2004 (first entry)

XX DE Drosophila melanogaster protein, SEQ ID 175.

XX KW Insecticide; Antiparasitic; Anthelmintic.

XX OS Drosophila melanogaster.

XX PN W02004039999-A2.

XX PD 13-MAY-2004.

XX PF 08-AUG-2003; 2003WO-US024982.

XX PR 30-OCT-2002; 2002US-0422377P.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Stam L, Kandar KP, Spana E, Bachmann J;

XX DR WPI; 2004-376203/35.

XX DR N-PSDB; ADS96553.

XX PT Identifying a compound that inhibits the activity of a protein for
PT Drosophila viability for use e.g., as insecticidal agent by expressing in
PT a recombinant host a DNA molecule to produce a protein essential for
PT Drosophila viability.

XX PS Claim 1; SEQ ID NO 175; 57pp; English.

XX CC The present invention relates to a method for identifying a compound that
XX inhibits the activity of a protein essential for Drosophila viability.
XX The method comprises: (a) expressing in a recombinant host a DNA sequence
XX encoding a protein essential for Drosophila viability; (b) testing
XX compounds suspected of having the ability to inhibit the activity of the
XX protein expressed in (a); and identifying a compound tested in (b) that
XX inhibits the activity of the protein. The method is useful in identifying
XX a compound that inhibits the activity of a protein essential for
XX Drosophila viability for use as insecticidal, ectoparasiticide,
XX antiparasitic, anthelmintic or acaricidal agent. The present sequence
XX is one such protein essential for Drosophila viability. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1525 AA;

Query Match 40.2%; Score 3128.5; DB 8; Length 1525;

Best Local Similarity 41.0%; Pred. No. 6e-278;

Matches 655; Conservative 303; Mismatches 458; Indels 183; Gaps 32;

QY 1 MAGNSLVPLVLMGRKAPTHICISAVLLTDDGATVIVGCHDQGLCLWDLS-VELQINPRAL 59

Db 1 MWTNLVVPVLVLMGPATPHICISVFLSDQFLTVTGCDQGLCLWQVEPTLKMSPRCL 60

QY 60 LFQHTASITCLSKACASSDKQYIVSABSGEMCLWDVSGRCIEPTKLCATHTGQFYQF 119

Db 61 LVGHSAPVLCVLRASLLPENNPLVSSSNGEMCTDLTDGKWEAVKLPQVHTQIQSYH- 119

QY 120 SVGNQREGLLCHGHTEILVVDATSLVLYSLVSKISPDWISSMGIIRSHRTQEDTVA 179

Db 120 -TANSEVDRLFCIGYTAIEVMVDPFSLVLYVLLSSKVPDWISAIHVLPRMRKDDVLA 178

QY 180 LSVTGILKWIIVTSEISDMQDTEPIFEBSKPIYQNCOSISFCATQBSLLVVCISKYR 239
Db 179 ITTGTGVKVTLTG--NENKHAEPYENESKEIRCLNALTMCNCAQNTQTLVLCVTKYQ 236
QY 240 VEDAGDYSLLCSPSENGOTWTGGDFVSSDKVITWTENGOSYIYKLPASCLPASDSRSD 299
Db 237 IYDAGDFTVLCSVIAPARERWQGDFTSDRVMLTDEGKYIYKLPANCIPOKNGFHS- 295
QY 300 VGKAVENLIPPVQHILLDRKDELLICPPVTRFFYGCIFYFHKLLIQGSSSGRLNIW--- 356
Db 296 --KSVVVDAPLYLYVLOHAGDK-VLSCPPAMKLLQAGGQHN--LARGDSEGIYVNVNP 350
QY 357 -----NISDTADKQSGSEGLAMTTSISLOAFDKLNPAGIIDQISVIPNSNEPKVTA 411
Db 351 EYPLDNISILQAKQMPRLPKHVCTSLVEASIMDPPVGIILDLQSLRITES--PVKLTLS 408
QY 412 SVYIPAHGRLVCGREDGSIIVPATQTAIVOLLQGEHMLRRGPPHRTLRGHNKTCCL 471
Db 409 SIYLPQOSRLVIGREDGSIIVPATQTVMMQLLVGIKQNFSDMPFHOIILYGHGRVNCCL 468
QY 472 YPHOVSAFYDQRYLISGVDVFSVIWIDIFSGEMKHIFCVHGGBITOLLVPPENC SARVOH 531
Db 469 CFSMTHSRYEKSHLLSGGIDPAVCLMDLYSGSLHFRFCVHAGEITOLLVPPESCSPRLK 528
QY 532 CICSVASDHSVGLLSLREKKICMLASRHLFPIQVIKWRPDDYLVVVGCSGSIYVWQMDT 591
Db 529 CICSVASDHSVTLVSLQERKCVTLASRHLFPVVTIKWRPLDDFLIVGCSGSIYVWQMET 588
QY 592 GAIIDRCVMGITAVEILNACDE-----AVPAAVDSLSPHAPVNLKQAMTRSL 637
Db 589 GHLDRLVHGLMABEVLSDACDEQADGGSGGSGNGASASEMGMANPAVHFFRGLKSRNM 648
QY 638 AALKWMAHKLQTLATNLLASEASDKGN--LPKYSNLSLMQAIKTNLTDPDIHVLFPD- 694
Db 649 NAIR-----NATRGITQLQOLQNGNDFLWKHRSNPLVQLGRINPKDAESHILFPD 704
QY 695 VEALIIQLITEASRNTALISPENLQKASGSDGSELTGKRAAVLFQOVK----- 747
Db 705 IEGLIFFELHSEVAQMTPATLESGLVHLQNPDKGSKMHLDAKCKIGDFFNKVKNKAVDVE 764
QY 748 EYIKENIKHLLDDEEDEIMR-----ORRESDPEYRSKSKPLTLLENLTM 797
Db 765 KILQKDKHGLVQKPKFKEIYEKKVQAKVESLQKAVEPEEQDLKSIASRMEYTHVM 824
QY 798 DTAKLFMSCLAHAGLNEVLDEVLDRGLMKHCTVSFGLISRGHMSLMPLQYNQ---- 853
Db 825 EVAQLLSLHSLWGLDPLHDKMCETRLGLRPLVPSYGVLSKAGTMSLLPTWQNNYAI 884
QY 854 -PACKL--SHGKTEVGRKLPASEGV-----GKGTGVSRVAVTTOHLLSIISLANTLMSMT 905
Db 885 PFGIQLPSSSKKRPPEELQRLHLEHTAVFTSRHLWELSTLTITNHLALVAMSNLLSMS 944
QY 906 NATFTGDHMKGPTRPPRSTPDLSKARGSPPTSSNIVOGQIKQVAAPVVSARSADHSG 965
Db 945 AASFL-----PDSEKHK-----KLQRLA-----QRTDSTLSN 971
QY 966 SDPPSAPALHTCFVNEGWSQLAAMHCVMLPD---LLGLDKRPPPLEMLARWDRCLSE 1022
Db 972 EBERBELMAHHSIQIKHAWSLATHCHFLPLPKIEALEPKPKRPOVEMVVRKHQHCIE 1031
QY 1023 VREAQALLLAEIRIEQAGRKEADAWAPYLPQYIDHVISPGVTSEAAQTITAP---- 1078
Db 1032 IREAAQILLGLTMRGKGRQLVESWAQYLPYLYTH--TEPIVGAQQOQALALISQASGG 1089
QY 1079 -----DASGPEAKV-----QEEHDLVDDDTTTCGLSVPQM 1110
Db 1090 AGSGSGNGGVGVGSGGAGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 1139
QY 1111 KKISYSYERRRQATAIVLLGVIGAEFGAEI--EPPK-----LLTRPRS 1152
Db 1140 RKPSSLSLSELRKQTTAVILLGVIGAEFGQDISQESPNHRGSIEMATGANLTSGVAGERR 1199

QY 1153 SSOIPGEGFLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPHTIRRTAIDLIGRFTVM 1212
Db 1200 KSSVVEFGIAN-----NLRLTSMALAHLLYAPPSPKLPQYTPLRRAIDLIGRFTVM 1254
QY 1213 EPTMYSVAVLWGELLECADEAKOLANITMGLPLSPAADSARSARHALSLIATARPFAFIT 1272
Db 1255 EPLYDVSXVLLGLEISCEG-KAVPNLYKPLTPQADACRTARHALRLIATARPAFIT 1313
QY 1273 TIAKEVHRHTALAAQTOSQONMIT-TTLARAKGEILRVIELIEKMPDVDLLVEMDI 1331
Db 1314 TMAREVARNTMOQNAQSINTPLTQSVLHKAKGEILQCCEMLIDKMQSEIAGLLVEMDI 1373
QY 1332 IMYCLEGLVKKGLOECRPAICRFYMWVSYYENHRIAVCARHGSVALYDIRTKCQOTIH 1391
Db 1374 ALHCVDNELKNGLAELCPAICKFNQISCAQTRRIAVGANGNLAIYELRONKQCMIP 1433
QY 1392 GHKGPITAVAPADGRYLATYSNTDGHISFWQNTSLLGSIMLNSAPQLRCIKTYQVPP 1451
Db 1434 ANTHPTITSLAFSPDGKYLVSYSCAENRLSPWQSTGMFG-LGQ-----SQTRCTKGYSTAP 1488
QY 1452 VQPASPGSHNALKLARLIWTSNRRNVILMAHDGKEHRPMV 1490
Db 1489 IPDVS--RLNPMRLAKLVWNNRTVTMLADGSETRFNV 1525

RESULT 8

ADA55388
ID ADA55388 standard; protein; 559 AA.

XX AC ADA55388;

XX DT 20-NOV-2003 (first entry)

XX DE Human protein, SEQ ID 2956.

XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.

XX OS Homo sapiens.

XX PN EP1293569-A2.

XX PD 19-MAR-2003.

XX PF 21-MAR-2002; 2002EP-00006586.

XX PR 14-SEP-2001; 2001JP-00328381.

XX PR 24-JAN-2002; 2002US-0350435P.

XX XX (HELI-) HELIX RES INST.

XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-395539/38.
DR N-PSDB; ADA53749.

XX XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.

XX XX Claim 14; SEQ ID NO 2956; 205pp; English.

XX XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.

XX XX Sequence 559 AA;

Query Match 36.1%; Score 2805.5; DB 6; Length 559;
Best Local Similarity 94.3%; Pred. No. 6.3e-249; Indels 33; Gaps 1;
Matches 549; Conservative 0; Mismatches 0;
QY 526 SARVQHICISVASDHSVGLLSLREKKCIMLASRHLFPQIVIKWRPDDLYLVVCGSDGSVY 585
Db 11 SARVQHICISVASDHSVGLLSLREKKCIMLASRHLFPQIVIKWRPDDLYLVVCGSDGSVY 70
QY 586 VQMDTGALDRCVMGITAVEILNACDEAVPAADVSLSHPAVNLKQAMTFRSLAALKNMAH 645
Db 71 VQMDTGALDRCVMGITAVEILNACDEAVPAADVSLSHPAVNLKQAMTFRSLAALKNMAH 130
QY 646 HKLQTLATNLWLASASDKNLPRYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTE 705
Db 131 HKLQTLATNLWLASASDKNLPRYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTE 190
QY 706 EASRPNTALISPENLQKASGSSDKGGSFLTGRKAAVLFOQVKETIKENIKEHLDDDEED 765
Db 191 EASRPNTALISPENLQKASGSSDKGGSFLTGRKAAVLFOQVKETIKENIKEHLDDDEED 250
QY 766 BEIMRQRRESDPEYRSSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRIG 825
Db 251 BEIMRQRRESDPEYRSSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRIG 310
QY 826 MLKPHCTVSFGLLSRGHMSLMLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVRSRA 885
Db 311 MLKPHCTVSFGLLSRGHMSLMLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVRSRA 370
QY 886 VTTQHLLSIISLANTIMSMNTATFIGDHMKGPTRPPRSTPDLKARGSPPTSSNIVQG 945
Db 371 VTTQHLLSIISLANTIMSMNTATFIGDHMKGPTRPPRSTPDLKARGSPPTSSNIVQG 430
QY 946 QIKQVAAFPVVSARSADHSGSDPPSPAPALHTCFLVNEGHSQLAAMHCVMPLDILGLDKFR 1005
Db 431 QIKQ-----GWSQLAAAHVCVMPLDILGLDKFR 457
QY 1006 PPLLEMLARRWDRCLVREAAQAALLLAELRRIEQAQRKEAIDAWAPYLPQYIDHVISPG 1065
Db 458 PPLLEMLARRWDRCLVREAAQAALLLAELRRIEQAQRKEAIDAWAPYLPQYIDHVISPG 517
QY 1066 VTSEAAQTITTAPDASGPEAKVQEEHDLVDDDDITTGCLSSV 1107
Db 518 VTSEAAQTITTAPDASGPEAKVQEEHDLVDDDDITTGCLSSV 559

RESULT 9

ADM04809
ID ADM04809 standard; protein; 775 AA.

XX AC ADM04809;

XX DT 20-MAY-2004 (first entry)

XX DE Human protein of the invention SEQ ID NO:3494.

XX KW human; gene therapy; diagnostic marker; pharmaceutical.

XX OS Homo sapiens.

XX PN EP1347046-A1.

XX PD 24-SEP-2003.

XX PF 12-APR-2002; 2002EP-00008400.

XX PR 22-MAR-2002; 2002JP-00137785.

XX XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231124P.
PR 08-SEP-2000; 2000US-0231124P.
PR 08-SEP-2000; 2000US-0231124P.
PR 08-SEP-2000; 2000US-0231133P.
PR 08-SEP-2000; 2000US-0231141P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231368P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241121P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.

PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476159/51.

N-P5DB; ADM19384.

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 11; SEQ ID NO 670; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a these. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. This sequence corresponds to a protein of the invention.

Sequence 209 AA;

Query Match 12.9%; Score 1003; DB 4; Length 209;

Best Local Similarity 98.5%; Pred. No. 2.8e-83;

Matches 191; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAGNSLVLPFVLWGRKAPTHCSAVLLTDDGATVTVGCHDQICLMDLSVFEIQINPRALL 60
|||||

Db 11 MAGNSLVLPIVLWGRKAPTHCISAVLTLDDGATVTCCHDQICLMDLSVELQINPRALL 70
Qy 61 FGHASITCLSKACASSDKQYIVSASSEGWCLWVSDGRCIETFKLACTHTTGIFQFYQFS 120
Db 71 FGHASITCLSKACASSDKQYIVSASSEGWCLWVSDGRCIETFKLACTHTTGIFQFYQFS 130
Qy 121 VGNOREGRLLCHGHYPILVVDATSLVLYSLVSKISPDWISSMIIIRSHRTQEDTIVVAL 180
Db 131 VGNOREGRLLCHGHYPILVVDATSLVLYSLVSKISPDWISSMIIIRSHRTQEDTIVVAL 190
Qy 181 SVTGILKVMWIVTSE 194
Db 191 SVTGILKVMWIVTSE 204

RESULT 11
ADM19833
ID ADM19833 standard; protein; 210 AA.
XX
AC ADM19833;
DT
DT 20-MAY-2004 (first entry)
DE
DE Protein encoded by novel human channel/transporter gene #151.
XX
XX immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytotatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW gene therapy; channel/transporter protein; rheumatoid arthritis;
KW neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.
XX
OS Homo sapiens.
XX
XX WO200154472-A2.
PN
PN 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001307.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214986P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228242P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244674P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476159/51.
XX N-PSDB; ADMI9354.

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 11; SEQ ID NO 640; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a these. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. This sequence corresponds to a protein of the invention.

XX Sequence 210 AA;

Query Match 11.7%; Score 909; DB 4; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.4e-74;

	Matches 173;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1302	AKGEILRVIEILIEKPTDVLVVDLLVEVMDIIMYCLEGSLVKKKGLQECPPAICRFYVVS	Y 1361		
DB	1	AKGEILRVIEILIEKPTDVLVVDLLVEVMDIIMYCLEGSLVKKKGLQECPPAICRFYVVS	Y 60		
QY	1362	YERNHRIAVGARHGSVALYDITRGKQTIHGKGPITAVAFAPDGRYLATYNTDHSISF	Y 1421		
DB	61	YERNHRIAVGARHGSVALYDITRGKQTIHGKGPITAVAFAPDGRYLATYNTDHSISF	Y 120		
QY	1422	WQNTSLLGSGIGMNSAPOLRCIKTYQVPPVQSPASPGSHNALKLARLIWTSNR	Y 1474		
DB	121	WQNTSLLGSGIGMNSAPOLRCIKTYQVPPVQSPASPGSHNALKLARLIWTSNR	Y 173		
RESULT 12					
ADM20106					
ID	ADM20106	standard; protein; 81 AA.			
XX	ADM20106;				
XX	20-MAY-2004	(first entry)			
DE		Protein encoded by novel human channel/transporter gene #181 clone 2.			
KW		immunosuppressive; antiarthritic; antirheumatic; antiproliferative;			
KW		cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;			
KW		neuroprotective; antibacterial; virucide; fungicide; ophthalmological;			
KW		gene therapy; channel/transporter protein; rheumatoid arthritis;			
KW		neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;			
KW		angiogenesis; nervous system disorder; Alzheimer's disease;			
KW		ocular disorder; corneal infection; wound healing;			
KW		epithelial cell proliferation; skin aging; sunburn; transplantation;			
KW		chemotaxis; food additive.			
XX					
OS		Homo sapiens.			
XX					
PN		WO200154472-A2.			
XX					
PD		02-AUG-2001.			
XX					
PF		17-JAN-2001; 2001WO-US001307.			
XX					
PR		31-JAN-2000; 2000US-0179065P.			
PR		04-FEB-2000; 2000US-0180628P.			
PR		24-FEB-2000; 2000US-0184664P.			
PR		02-MAR-2000; 2000US-0186350P.			
PR		16-MAR-2000; 2000US-0189874P.			
PR		17-MAR-2000; 2000US-0190076P.			
PR		18-APR-2000; 2000US-0198123P.			
PR		19-MAY-2000; 2000US-0205515P.			
PR		07-JUN-2000; 2000US-0209467P.			
PR		28-JUN-2000; 2000US-0211486P.			
PR		30-JUN-2000; 2000US-0215135P.			
PR		07-JUL-2000; 2000US-0216647P.			
PR		07-JUL-2000; 2000US-0216880P.			
PR		11-JUL-2000; 2000US-0217487P.			
PR		11-JUL-2000; 2000US-0217496P.			
PR		14-JUL-2000; 2000US-0218290P.			
PR		26-JUL-2000; 2000US-0220963P.			
PR		26-JUL-2000; 2000US-0220964P.			
PR		14-AUG-2000; 2000US-0224518P.			
PR		14-AUG-2000; 2000US-0224519P.			
PR		14-AUG-2000; 2000US-0225213P.			
PR		14-AUG-2000; 2000US-0225214P.			
PR		14-AUG-2000; 2000US-0225266P.			
PR		14-AUG-2000; 2000US-0225267P.			
PR		14-AUG-2000; 2000US-0225268P.			
PR		14-AUG-2000; 2000US-0225270P.			
PR		14-AUG-2000; 2000US-0225447P.			
PR		14-AUG-2000; 2000US-0225757P.			
PR		14-AUG-2000; 2000US-0225758P.			
PR		14-AUG-2000; 2000US-0225759P.			

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:12:11 ; Search time 215 Seconds
(without alignments)
3548.828 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773

Sequence: 1 MAGNSLVLPVLWGRKAPTH.....TSNRNVLMAHDGKEHRFV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7773	100.0	1490	2 Q86UX5	Q86UX5 homo sapien
2	7573.5	97.4	1457	2 Q96P87	Q96P87 homo sapien
3	7408.5	95.3	1489	2 Q92019	Q92019 mus musculus
4	7373	94.9	1488	2 Q9ERH3	Q9ERH3 rattus norv
5	7254.5	93.3	1457	2 Q92070	Q92070 mus musculus
6	7247.5	93.2	1457	2 Q92018	Q92018 rattus norv
7	6023	77.5	1160	1 WDR7 HUMAN	Q9Y4E6 homo sapien
8	5784.5	74.4	1166	2 Q80TY3	Q80TY3 mus musculus
9	4331	55.7	956	2 Q6GPQ7	Q6GPQ7 xenopus lae
10	3170	40.8	614	2 Q86VP2	Q86VP2 homo sapien
11	3128.5	40.2	1525	2 Q9W543	Q9W543 drosophila
12	3107.5	40.0	1473	2 Q7Q3S0	Q7Q3S0 anopheles g
13	2944	37.9	1471	2 Q46092	Q46092 drosophila
14	2667	34.3	531	2 Q8C711	Q8C711 mus musculus
15	2006.5	25.8	1335	2 Q9BL05	Q9BL05 caenorhabdi
16	1610	20.7	1102	2 Q7Z313	Q7Z313 homo sapien
17	1442	18.6	775	2 Q8N8X2	Q8N8X2 homo sapien
18	839	10.8	343	2 Q960M9	Q960M9 drosophila
19	798	10.3	1342	2 Q86L37	Q86L37 dictyosteli
20	518.5	6.7	1446	2 Q9SUT4	Q9SUT4 arabisdopsis
21	331	4.3	117	2 Q8BPW7	Q8BPW7 mus musculus
22	234	3.0	1526	1 YV46 ANASP	Q8YR11 anabaena sp
23	230.5	3.0	1227	2 Q8Z0R1	Q8Z0R1 anabaena sp
24	229	2.9	1258	1 YS00 ANASP	Q8YTC2 anabaena sp
25	223	2.9	1184	2 Q7ND85	Q7ND85 gloeobacter
26	221.5	2.8	1081	2 Q7NCT8	Q7NCT8 gloeobacter
27	217.5	2.8	1592	2 Q62471	Q62471 caenorhabdi
28	206.5	2.7	1923	2 Q6AWF2	Q6AWF2 drosophila
29	205.5	2.6	1193	2 Q7ND05	Q7ND05 gloeobacter
30	204	2.6	1680	2 Q9UAQ2	Q9UAQ2 caenorhabdi
31	203.5	2.6	1601	2 Q8W5M1	Q8W5M1 caenorhabdi

RESULT 1

ID	Q86UX5	PRELIMINARY;	PRT;	1490 AA.
AC	Q86UX5;			
DT	01-JUN-2003 (TrEMBLrel. 24, Created)			
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Rabconnectin-3 beta.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22671656; PubMed=12786944;			
RA	Kawabe H., Sakisaka T., Yasumi M., Shingai T., Izumi G., Nagano F.,			
RA	Deguchi-Tawarada M., Takeuchi M., Nakanishi H., Takai Y.,			
RT	"A novel rabconnectin-3-binding protein that directly binds a GDP/GTP			
RT	exchange protein for Rab3A small G protein implicated in Ca(2+)-			
RT	dependent exocytosis of neurotransmitter."			
RL	Genes Cells 8:537-546(2003).			
CC	-1- SIMILARITY: Contains 6 WD repeats.			
DR	EMBL; AY093325; AM33134.1; -			
DR	InterPro; IPR001680; WD40.			
DR	InterPro; IPR011046; WD40-like.			
DR	Pfam; PF00400; WD40; 6.			
DR	PRINTS; PR00320; GPROTEINRPT.			
DR	SMART; SM00320; WD40; 7.			
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.			
DR	PROSITE; PS50082; WD_REPEATS_2; 4.			
DR	PROSITE; PS50294; WD_REPEATS_REGION; 3.			
KW	Repeat; WD repeat.			
SQ	SEQUENCE 1490 AA; 163809 MW; 626E876245696A1P CRC64;			
Query Match 100.0%; Score 7773; DB 2; Length 1490;				
Best Local Similarity 100.0%; Pred. No. 0;				
Matches 1490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	MAGNSLVLPVLWGRKAPTHCSAVLLTDDGATIVTGDHGOICLWDLVELQINPRALL	60	
Db	1	MAGNSLVLPVLWGRKAPTHCSAVLLTDDGATIVTGDHGOICLWDLVELQINPRALL	60	
QY	61	FIGHTASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLTHTGTGQFYQFS	120	
Db	61	FIGHTASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLTHTGTGQFYQFS	120	
QY	121	VGNQREGRLCHGHYPEILVDATSLVLYSVLSKISPDWISSMSIIRSHRQEDTVVAL	180	
Db	121	VGNQREGRLCHGHYPEILVDATSLVLYSVLSKISPDWISSMSIIRSHRQEDTVVAL	180	
QY	181	SVTGILKWIWTVSEISDMQDTEPIFEESKPIYCONCOSISFCAPTORSLVVCYSKYRW	240	
Db	181	SVTGILKWIWTVSEISDMQDTEPIFEESKPIYCONCOSISFCAPTORSLVVCYSKYRW	240	

ALIGNMENTS

32	202	2.6	1683	1	YL24 ANASP	Q8YV57 anabaena sp
33	200.5	2.6	1671	2	Q7NNP0	Q7NNP0 gloeobacter
34	199.5	2.6	1541	2	Q8HXL3	Q8HXL3 sus scrofa
35	199.5	2.6	1701	2	Q9V8I6	Q9V8I6 drosophila
36	199	2.6	1328	2	Q9U4H2	Q9U4H2 drosophila
37	199	2.6	1698	2	Q9V8I5	Q9V8I5 drosophila
38	199	2.6	1922	2	Q8MLH7	Q8MLH7 drosophila
39	197.5	2.5	929	2	Q19433	Q19433 caenorhabdi
40	193	2.5	778	2	Q7Q938	Q7Q938 anopheles g
41	192	2.5	1049	2	Q9ZEM4	Q9ZEM4 streptomyce
42	191.5	2.5	1721	2	Q7NID9	Q7NID9 gloeobacter
43	189	2.4	494	2	Q18295	Q18295 caenorhabdi
44	188.5	2.4	522	2	Q6GPC6	Q6GPC6 xenopus lae
45	188.5	2.4	1197	2	Q7NJ67	Q7NJ67 gloeobacter

```
QY 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
Db 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIPPVQHILDRKDKELLICPPVTRFPYGCREFYFHKLLIQGDSGRLNIWNISD 360
Db 301 GRAVENLIPPVQHILDRKDKELLICPPVTRFPYGCREFYFHKLLIQGDSGRLNIWNISD 360
QY 361 TADKQSEGLAMTTSISLQEAFLKLPAGIIDLQSVIPNSNEPLKVTASVYIPAHGR 420
Db 361 TADKQSEGLAMTTSISLQEAFLKLPAGIIDLQSVIPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCGRDGSIVIVPAQTQTAIVQLLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
Db 421 LVCGRDGSIVIVPAQTQTAIVQLLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
QY 481 DQRYLISGGVDRSVIIWDIFSGEMKHFVHGGEITQLLVPPENCARSARVOHCICSVASDH 540
Db 481 DQRYLISGGVDRSVIIWDIFSGEMKHFVHGGEITQLLVPPENCARSARVOHCICSVASDH 540
QY 541 SVGLLSLRKKCIIMLASRHLFPLOVTKWRPSDDYLWVCGSDGSVYVQMDTGALDRCVNG 600
Db 541 SVGLLSLRKKCIIMLASRHLFPLOVTKWRPSDDYLWVCGSDGSVYVQMDTGALDRCVNG 600
QY 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSALALKNMAHHLQTLATNLILASEA 660
Db 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSALALKNMAHHLQTLATNLILASEA 660
QY 661 SDKGNLPKYSHNSLAWQAIKTNLTDPDIHVLFPDVEALIIQLLITEASRPNLTALISPENL 720
Db 661 SDKGNLPKYSHNSLAWQAIKTNLTDPDIHVLFPDVEALIIQLLITEASRPNLTALISPENL 720
QY 721 QKASGSDSGGSLTKRAAVLFOQVKETIKENIKHLLDDEEDEBEIMRQRRESDPEY 780
Db 721 QKASGSDSGGSLTKRAAVLFOQVKETIKENIKHLLDDEEDEBEIMRQRRESDPEY 780
QY 781 RSKSKPLTLLEVLNMTDAKLFMSCLHAWGLNEVLDEVLDRGLMKLPHCTVSGFLLSR 840
Db 781 RSKSKPLTLLEVLNMTDAKLFMSCLHAWGLNEVLDEVLDRGLMKLPHCTVSGFLLSR 840
QY 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGYGVSRAVTTQHLLSIIISLANT 900
Db 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGYGVSRAVTTQHLLSIIISLANT 900
QY 901 LMSMTNATIGDHMKKGPTRPPRSTPDLISKARGSPPTSSNIVQGOIKOVAAPVVSARS 960
Db 901 LMSMTNATIGDHMKKGPTRPPRSTPDLISKARGSPPTSSNIVQGOIKOVAAPVVSARS 960
QY 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDR 1020
Db 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDR 1020
QY 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPYLPQYIDHVTSPGVTSEAAQTITAPDA 1080
Db 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPYLPQYIDHVTSPGVTSEAAQTITAPDA 1080
QY 1081 SGPEAKVQEEHDLVDDDDITTCCLSSVPQMKKISTSYEERRKOATAIVLLGVIGAFEGAE 1140
Db 1081 SGPEAKVQEEHDLVDDDDITTCCLSSVPQMKKISTSYEERRKOATAIVLLGVIGAFEGAE 1140
QY 1141 IEPKLLTTPRSSQIPEGFGLTSGGSNTSLARHTCKALTFLLOPPSPKLPHPHSTIR 1200
Db 1141 IEPKLLTTPRSSQIPEGFGLTSGGSNTSLARHTCKALTFLLOPPSPKLPHPHSTIR 1200
QY 1201 AIDLIGRGFTVWPEYMDVSAVLMGLLELCADAEKQLANITMGLPLSPAADSARSARHALS 1260
Db 1201 AIDLIGRGFTVWPEYMDVSAVLMGLLELCADAEKQLANITMGLPLSPAADSARSARHALS 1260
QY 1261 LIATARPFAFITIAKEVHRHTALAAANTOSQOQNMHTTTLARAKGEILRVIELIEKMP 1320
Db 1261 LIATARPFAFITIAKEVHRHTALAAANTOSQOQNMHTTTLARAKGEILRVIELIEKMP 1320
QY 1321 VVDLLVEVMDIIMYCLEGSLVKKGLQECFPAICRFYVMVSYERNHRIAIVGARHGSVALY 1380
```

```
Db 1321 VVDLLVEVMDIIMYCLEGSLVKKGLQECFPAICRFYVMVSYERNHRIAIVGARHGSVALY 1380
QY 1381 DIRTGKQCTIHGKGPITAVAPADGRYLATVNTDSDSHISFQWMTSLGSGIMLNSAPQ 1440
Db 1381 DIRTGKQCTIHGKGPITAVAPADGRYLATVNTDSDSHISFQWMTSLGSGIMLNSAPQ 1440
QY 1441 LRCIKTYQVPVQPPASPGSHNALKARLIWTSNRNVILMAHDGKEHFRMV 1490
Db 1441 LRCIKTYQVPVQPPASPGSHNALKARLIWTSNRNVILMAHDGKEHFRMV 1490

RESULT 2
Q96PS7 PRELIMINARY; PRT; 1457 AA.
AC Q96PS7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE TGF-beta resistance-associated protein TRAG (Fragment).
GN Name=TRAG;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Sanders S., Thorgeirsson S.S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Contains 6 WD repeats.
EMBL: AF188125; AAL03983.1; -.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR011046; WD40_like.
DR Pfam: PF00400; WD40; 6.
DR PRINTS: PR00320; GPROTEINERPT.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS00082; WD_REPEATS_2; 4..
DR PROSITE: PS00294; WD_REPEATS_REGION; 3.
KW Repeat; WD repeat.
FT NON_TER 1457 1457
SQ SEQUENCE 1457 AA; 160494 MW; 202B6A1C5EB0E2F7 CRC64;

Query Match 97.4%; Score 7573.5; DB 2; Length 1457;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1456; Conservative 1; Mismatches 0; Indels 33; Gaps 1;

QY 1 MAGNSILVLPVLWGRKAPTHCISAVLLTDDGATIVTGCHDGOICLWDLSELQINPRALL 60
Db 1 MAGNSILVLPVLWGRKAPTHCISAVLLTDDGATIVTGCHDGOICLWDLSELQINPRALL 60
QY 61 FGHASITCLSKACASSDKQYIVSASEGEMCLWVSDGRCIEFTKLAHTHTGQIYQFS 120
Db 61 FGHASITCLSKACASSDKQYIVSASEGEMCLWVSDGRCIEFTKLAHTHTGQIYQFS 120
QY 121 VGNQREGRLCHGHYPEILVVDATSLVLYSVLSKISPDWISSMSTIRSHRTOEDTVVAL 180
Db 121 VGNQREGRLCHGHYPEILVVDATSLVLYSVLSKISPDWISSMSTIRSHRTOEDTVVAL 180
QY 181 SVTGILKVMIVTSEISDMQDTEPIFEESKPIYCONCOSISFCFTQRLVLLVCSKYWRV 240
Db 181 SVTGILKVMIVTSEISDMQDTEPIFEESKPIYCONCOSISFCFTQRLVLLVCSKYWRV 240
QY 241 FDAGYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
Db 241 FDAGYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIPPVQHILDRKDKELLICPPVTRFPYGCREFYFHKLLIQGDSGRLNIWNISD 360
Db 301 GRAVENLIPPVQHILDRKDKELLICPPVTRFPYGCREFYFHKLLIQGDSGRLNIWNISD 360
QY 361 TADKQSEGLAMTTSISLQEAFLKLPAGIIDLQSVIPNSNEPLKVTASVYIPAHGR 420
```

Db 361 TADKQSGSEGLAMTTSISLQAFDKNPCPAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCREDGSIIVIPATQTAIVOLLQGEHMLRRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
Db 421 LVCREDGSIIVIPATQTAIVOLLQGEHMLRRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
QY 481 DQRYLISGGVDVFSVIINDISFEGMKHIFCVHGEITQLLPPENCARSARVQHCICSVASDH 540
Db 481 DQRYLISGGVDVFSVIINDISFEGMKHIFCVHGEITQLLPPENCARSARVQHCICSVASDH 540
QY 541 SVGLLSUREKCKTCLASRHLFPQIVIKWRPDDVLLVVGCSGSGSVYVWMDTGALDRCVMG 600
Db 541 SVGLLSUREKCKTCLASRHLFPQIVIKWRPDDVLLVVGCSGSGSVYVWMDTGALDRCVMG 600
QY 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQAMTERRSLAALKNMAHKLQTLATNLLASEA 660
Db 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQAMTERRSLAALKNMAHKLQTLATNLLASEA 660
QY 661 SDKGNLPKYSHNSLMQAIKTNLTDPDIHVLFPDVEALIIQLLTHEASRNTALISPENL 720
Db 661 SDKGNLPKYSHNSLMQAIKTNLTDPDIHVLFPDVEALIIQLLTHEASRNTALISPENL 720
QY 721 QKASGSDKGSFELTGKRAAVLFOQVKETIKENIKHLLDDEEDEDEIMEQREESDPEY 780
Db 721 QKASGSDKGSFELTGKRAAVLFOQVKETIKENIKHLLDDEEDEDEIMEQREESDPEY 780
QY 781 RSSKSKPLTLLLEYNLTMDTAKLPMSCILHAWGLNEVLDEVCILRLGMLKPHCTVSGFLLSR 840
Db 781 RSSKSKPLTLLLEYNLTMDTAKLPMSCILHAWGLNEVLDEVCILRLGMLKPHCTVSGFLLSR 840
QY 841 GGHMSLMLPGYNOPACKLSHGKTEVGKRLPASBGVGKGTGVSRVAVTQHLISIIISLANT 900
Db 841 GGHMSLMLPGYNOPACKLSHGKTEVGKRLPASBGVGKGTGVSRVAVTQHLISIIISLANT 900
QY 901 LMSMTNATFIDGHWKKGTPRPPRSTPDLKARGSPPTSSNIYQGIKQVAAVPVARS 960
Db 901 LMSMTNATFIDGHWKKGTPRPPRSTPDLKARGSPPTSSNIYQGIKQVAAVPVARS 960
QY 961 ADHSGSDPPSPALHTCFLVNEGWSOLAAMHCVMPLDGLGLDFRPPLEMLARRWQDRC 1020
Db 961 ADHSGSDPPSPALHTCFLVNEGWSOLAAMHCVMPLDGLGLDFRPPLEMLARRWQDRC 1020
QY 1021 LEVREAAQALLAELRRIEQAQKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1080
Db 988 LEVREAAQALLAELRRIEQAQKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1047
QY 1081 SGPEAKVQEEHDLVDDDIITGCLSSVPQMKKISTSYEERRKQATAIVLLGVIGAEFGAE 1140
Db 1048 SGPEAKVQEEHDLVDDDIITGCLSSVPQMKKISTSYEERRKQATAIVLLGVIGAEFGAE 1107
QY 1141 IEPKLLTRPRSSSQIPEGGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPPhSTIRRT 1200
Db 1108 IEPKLLTRPRSSSQIPEGGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPPhSTIRRT 1167
QY 1201 AIDLIGRGFTWEPYMDVSAVLMGLLELCAEAKQOLANITMGLPLSPAADSARSARHALS 1260
Db 1168 AIDLIGRGFTWEPYMDVSAVLMGLLELCAEAKQOLANITMGLPLSPAADSARSARHALS 1227
QY 1261 LIATAPPAITTAKEVHRHTALANTOSQONMHTTLARAKEILRVTEILIEKMPDT 1320
Db 1228 LIATAPPAITTAKEVHRHTALANTOSQONMHTTLARAKEILRVTEILIEKMPDT 1287
QY 1321 VDDLVEVMDIIMVCLGSLVKKGLQECPPALCRFMYSYERNHRIAVGARHGSVALY 1380
Db 1288 VDDLVEVMDIIMVCLGSLVKKGLQECPPALCRFMYSYERNHRIAVGARHGSVALY 1347
QY 1381 DIRTGKQTHGHKGPITAVAPADGRYLATYNTDISHISFWQMTSLGSGIMLNSAPQ 1440
Db 1348 DIRTGKQTHGHKGPITAVAPADGRYLATYNTDISHISFWQMTSLGSGIMLNSAPQ 1407
QY 1441 LRCIKTYQVPPVQSPGSHNALKLARLIWTSNRNVILMAHDKGHRFMV 1490

Db 1408 LRCIKTYQVPPVQSPGSHNALKLARLIWTSNRNVILMAHDKGHRFMV 1457
RESULT 3
Q92019
ID Q92019 PRELIMINARY; PRT; 1489 AA.
AC Q92019
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE TGF-beta resistance-associated protein TRAG (Fragment).
GN Name=wdr7; Synonyms=Trag;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ; TISSUE=Brain;
RA Sanders S.; Thorgeirsson S.S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 6 WD repeats.
DR EMBL; AF188124; AAL03982.1; -
DR MGD; MGI:1860197; Wdr7.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE; PS50082; WD_REPEATS_2; 4.
DR PROSITE; PS50294; WD_REPEATS_REGION; 3.
KW Repeat; WD repeat.
FT NON TER 1489
SQ SEQUENCE 1489 AA; 163387 MW; 6B50C7958F0D7940 CRC64;
Query Match 95.3%; Score 7408.5; DB 2; Length 1489;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 1414; Conservative 36; Mismatches 39; Indels 1; Gaps 1;
QY 1 MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATVITGCHDGOICLWLSVLELQINPRALL 60
Db 1 MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATVITGCHDGOICLWLSVLELQINPRALL 60
QY 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGIGQYQFS 120
Db 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGIGQYQFS 120
QY 121 VGNQREGRLCHGHYPEILLVVDATSLVLYSLVSKISPDWISSMIIIRSHRTQEDTVVAL 180
Db 121 VGNQREGRLCHGHYPEILLVVDATSLVLYSLVSKISPDWISSMIIIRSHRTQEDTVVAL 180
QY 181 SVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQSIISFCFTQSLVVCSSKYRW 240
Db 181 SVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQSIISFCFTQSLVVCSSKYRW 240
QY 241 FDAGDYSLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
Db 241 FDAGDYSLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIIPVQHILLDRKOKELLICPPVTRFPYGCREFYHKLIIQDSSGRINIWNISD 360
Db 301 GRAVENLIIPVQHILLDRKOKELLICPPVTRFPYGCREFYHKLIIQDSSGRINIWNISD 360
QY 361 TADKQSGSEGLAMTTSISLQAFDKNPCPAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420
Db 361 TADKQSGSEGLAMTTSISLQAFDKNPCPAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCREDGSIIVIPATQTAIVOLLQGEHMLRRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
Db 421 LVCREDGSIIVIPATQTAIVOLLQGEHMLRRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
QY 481 DQRYLISGGVDVFSVIINDISFEGMKHIFCVHGEITQLLPPENCARSARVQHCICSVASDH 540
Db 481 DQRYLISGGVDVFSVIINDISFEGMKHIFCVHGEITQLLPPENCARSARVQHCICSVASDH 540

661 SDKGNLPKYSNLSMVAQIKTNLTDPDIHVLFFDVEALIIQLLTERASRNTALISPENL 720
721 QKASGSDKGSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
721 QKASGSDKGSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
781 RSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDVCLDRLGMLKPHCTVSGFLLSR 840
781 RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDVCLDRLGMLKPHCTVSGFLLSR 840
841 GGHMSLWPCYNQPACKLSHGKTEVGRKLPASGVGKGTGVSRAVTTQHLISILANT 900
841 GGHMSLWPCYNQAAAGKLLQAKAAGRGKGPATSGVGTGTVSRAVTTQHLISILANT 900
901 LMSMTNATFGIDHMKGTPTPRPSTPDLISKARGSPPTSSNIVQGOIKQVAAPVWSARSD 960
901 LMSMTNATFGIDHMKGTPTPRPSTPDLISKARGSPPTSSNIVQGOIKQVAAPVWSARSD 960
961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAMCHVMLPDLGLDPRPPLLEMLARRWQDRC 1020
960 ADHSGS-ASASPALRTCFVNEGWSQLAAMCHVMLPDLGLDPRPPLLEMLARRWQDRC 1018
1021 LEVREAAQALLAELRIEAGKEAIDAWAPYLPOYIDHVISPGVTSEAAQTITAPDA 1080
1019 LEVREAAQALLAELRIEAGKETIDTWPALPQMDHVISPGVTAEAMQTMAAAPDA 1078
1081 SGPEAKVQEEHDLVDDITGCLSSVPQMKKISTSYEERKQATAVILGVTAEGFAG 1140
1079 SGPEAKVQEEHDLVDDITGCLSSVPQMKKISTSYEERKQATAVILGVTAEGFAG 1138
1141 IEPPKLLTRPRSSQIPEGFLTSGGSNSYLARHTCKALFFLLQPPSPKLPHPSTIRRT 1200
1139 IEPPKLLTRPRSSQIPEGFLTSGGSNSYLARHTCKALFFLLQPPSPKLPHPSTIRRT 1198
1201 AIDLIGRFTVWEPYMDVSAVLMGLLELCADEKOLANTMGLPLSPAADSARSARHALS 1260
1199 AIDLIGRFTVWEPYMDVSAVLMGLLELCADEKOLANTMGLPLSPAADSARSARHALS 1258
1261 LIATAPPAFTTIAKEVHRHTALANTQSQNMHTTTLARAKGEILRVIEILIEKMPDT 1320
1259 LIATAPPAFTTIAKEVHRHTALANTQSQNHITTLARAKGEILRVIEILIEKMPDT 1318
1321 VDLLEVMVDIIMYCLEGLSVKKGLEQCPAICRFYMWVSYERNHRIAVGARHGSVALY 1380
1319 VDLLEVMVDIIMYCLEGLSVKKGLEQCPAICRFYMWVSYERNHRIAVGARHGSVALY 1378
1381 DIRTGKQTHGKGPITAVAPADGRYLATYNTDHSISFWQNTSLGSIOMLNSAPQ 1440
1379 DIRTGKQTHGKGPITAVSPADGRYLATYNTDHSISFWQNTSLGSIOMLNSAPQ 1438
1441 LRCIKTYQVPVOPASPGSHNALKLARLIWTSNRNVLMAHDKGHEFRW 1490
1439 LRCIKTYQVPVOPASPGSHNALKLARLIWTSNRNVLMAHDKGHEFRW 1488

RESULT 5
Q920J0

ID Q920J0 PRELIMINARY; PRT; 1457 AA.
AC Q920J0;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE TGF-beta resistance-associated protein TRAG (Fragment).
GN Name=Wdr7; Synonyms=Trag;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ; TISSUE=Liver;
RA Sanders S., Thorpeirsson S.S.;

Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -I- Similarity: Contains 6 WD repeats.
DR EMBL; AF186123; AAL03981.1; -.
DR MGD; MGI:1860197; wdr7.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINERPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE; PS0082; WD_REPEATS_2; 4.
DR PROSITE; PS0294; WD_REPEATS_REGION; 3.
KW Repeat; WD repeat.
FT NON_TER 1457 1457
SQ SEQUENCE 1457 AA; 160262 MW; EF33D0E44869E60A CRC64;

Query Match 93.3%; Score 7254.5; DB 2; Length 1457;
Best Local Similarity 93.0%; Pred.No. 0;
Matches 1386; Conservative 36; Mismatches 35; Indels 33; Gaps 1;

QY 1 MAGNSLVPLVILWGRKAPTHCISAVLLTDDGATIVTGDGQICLWDLSVELQINPRALL 60
DB 1 MAGNSLVPLVILWGRKAPTHCISAVLLTDDGATIVTGDGQICLWDLSVELQINPRALL 60
QY 61 FGHASITCLSKACASSDKOYIVSASESGEMCLWDVSDGRCIEFTKCLACTHTGTGQYQFS 120
DB 61 FGHASITCLSKACASSDKOYIVSASESGEMCLWDVSDGRCIEFTKCLACTHTGTGQYQFS 120
QY 121 VGNQREGRLLCHGHYPEILVVDATSLVLSVSKISPDWISSMSIIRSHRTQEDTVVAL 180
DB 121 VGNQREGRLLCHGHYPEILVVDATSLVLSVSKISPDWISSMSIIRSHRTQEDTVVAL 180
QY 181 SVTGILKWIWTSBISDMQDTBPIFEESKPIYQNCQISFCAPTORSLVVCVKYRW 240
DB 181 SVTGILKWIWTSBISDMQDTBPIFEESKPIYQNCQISFCAPTORSLVVCVKYRW 240
QY 241 FQAGDYSLSCSPSENGQWTGDFVSSDKVLIWTEGOSYIKLPASCLPASDSFRSDV 300
DB 241 FQAGDYSLSCSPSENGQWTGDFVSSDKVLIWTEGOSYIKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIIPVQHILDRDKKELLICPPVTRFPYGCYREYFHKLLIOGDSGRLNIWNISD 360
DB 301 GRAVENLIIPVQHILDRDKKELLICPPVTRFPYGCYREYFHKLLIOGDSGRLNIWNISD 360
QY 361 TADKQSEGLAWTTSISLOEAFDKLNPAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420
DB 361 TADKQSEGLAWTTSISLOEAFDKLNPAGIIDQSVIPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCGREDGSIIVPATOTAIVOLLOGEHMLRGWPHRTLRGRNKVTCLLYPHVSARY 480
DB 421 LVCGREDGSIIVPATOTAIVOLLOGEHMLRGWPHRTLRGRNKVTCLLYPHVSARY 480
QY 481 DORYLISGVDFSVIITWIDIFSGEMKHI FCHVGGEITQLLVPPENC SARVQHICISVASDH 540
DB 481 DORYLISGVDFSVIITWIDIFSGEMKHI FCHVGGEITQLLVPPENC SARVQHICISVASDH 540
QY 541 SVGLLSLRKKKCIIMLASRHLFPQIKWRPDDYLVVCGSDGSVYVQMDTGALDRCYMG 600
DB 541 SVGLLSLRKKKCIIMLASRHLFPQIKWRPDDYLVVCGSDGSVYVQMDTGALDRCYMG 600
QY 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQMTTRSLAALKMAHKLQTLATNLASEA 660
DB 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQMTTRSLAALKMAHKLQTLATNLASEA 660
QY 661 SKGNLPKYSNLSMVAQIKTNLTDPDIHVLFFDVEALIIQLLTERASRNTALISPENL 720
DB 661 SKGNLPKYSNLSMVAQIKTNLTDPDIHVLFFDVEALIIQLLTERASRNTALISPENL 720
QY 721 QKASGSDKGSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
DB 721 QKASGSDKGSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
QY 781 RSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDVCLDRLGMLKPHCTVSGFLLSR 840
DB 781 RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDVCLDRLGMLKPHCTVSGFLLSR 840

QY	841	GGHMSLMLPYNOPACKLSHGKTEVGRKLPASBGVKGTYGVSRAVTTQHLISISLANT	900
Db	841	GGHMSLMLPYNOAAGKLHAKAEVGRKLPABEGVGKGYTVSRAVTTQHLISISLANT	900
QY	901	LMSMTNATFIGDHMKKGTPRPPRPTDLSKARGSPPTSSNIVQGIQKQVAAPVVSARS	960
Db	901	LMSMTNATFIGDHMKKGTPRPPRPTDLSKARDSPPASSNIVQGIQKQ-----	949
QY	961	ADHSGSDPPSAPALHTCFVLNKGWSQLAAHMCVMLPDLGLGDKFRPPLLEMLARRWQDRC	1020
Db	950	-----GWSQLAAHMCVMLPDLGLGLERFRPPLLEMLARRWQDRC	987
QY	1021	LEVREAAQALLAEARRIEQAGKEAIDAWAPVLPQYIDHVISPGVTSAAQTITTPADA	1080
Db	988	LEVREAAQALLAEARRIEQAGKETIDWAPVLPQYMDHVISPGVTSAGAMTMAAPDA	1047
QY	1081	SGPEAKVQBEEDHLVDDDDITTCGLSSVPQMKKISTSYEERRKQATAIVLLGVIGAEFGAE	1140
Db	1048	SGPEAKVQBEEDHLVDDDDITAGCLSSVPQMKKISTSYEERRKQATAIVLLGVIGAEFGAE	1107
QY	1141	IEPPKLLTPRSSSQIPEGFGLTSGGSNTSLARHTCKALTILLQPPSPKLPHPSTIRRT	1200
Db	1108	IEPPKLLTPRSSSQIPEGFGLTSGGSNTSLARHTCKALTYLLQPPSPKLPHPSTIRRT	1167
QY	1201	AIDLIGRGFTVPEPYMDVSAVLMGLLELCADAEKOLANTMGLPLSPAADSARSARHALS	1260
Db	1168	ATDLIGRGFTVPEPYMDVSAVLMGLLELCADAEKOLANTMGLPLSPAADSARSARHALS	1227
QY	1261	LIATARPPAFITIAKEVHRHTALAANTOSQNMHTTTLARAKGETLRVIELIEKMPDT	1320
Db	1228	LIATARPPAFITIAKEVHRHTALAANTOSQSIHTTTLARAKGETLRVIELIEKMPDT	1287
QY	1321	VDLLVEVMDIIMYCLEGSLVKKGLQECFPAICRFYMSYVERNHRIAVGARHGSVALY	1380
Db	1288	VDLLVEVMDIIMYCLEGSLVKKGLQECFPAICRFYMSYVERSHRIAVGARHGSVALY	1347
QY	1381	DIRTGKQCTHGHKGPITAVAPADGRYLATYSNTDISHISFWMQNTSLGSGTMLNSAPQ	1440
Db	1348	DIRTGKQCTHGHKGPITAVSPADGRYLATYSNTDISHISFWMQNTSLGSGTMLNSAPQ	1407
QY	1441	LRCIKTYQVPPVQVAPSPGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV	1490
Db	1408	LRCIKTYQVPPVQVAPSPGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV	1457
RESULT 6			
ID	Q92018	PRELIMINARY; PRT; 1457 AA.	
AC	Q92018;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	TCF-beta resistance-associated protein TRAG (Fragment).		
GN	Name=Trag;		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	Sanders S., Thorgeirsson S.S.;		
RL	Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.		
CC	-I- SIMILARITY: Contains 6 WD repeats.		
CC	EMBL; AF192379; AAL03984.1; -.		
DR	Pfam; PF00400; WD40; 6.		
DR	PRINTS; PR00320; GPOTEINERPT.		
DR	SMART; SM00320; WD40; 6.		
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.		
DR	PROSITE; PS50082; WD_REPEATS_2; 4.		
DR	PROSITE; PS50294; WD_REPEATS_REGION; 3.		
KW	Repeat; WD repeat.		

FT	NON TER	1457	1457
SQ	SEQUENCE	1457 AA;	160192 MW; 5E4AEAL3BCFF3B2 CRC64;
Query Match			
Best Local Similarity 93.2%; Score 7247.5; DB 2; Length 1457;			
Matches 1384; Conservative 37; Mismatches 36; Indels 33; Gaps 1;			
QY	1	MAGNSILPVLWGKAPTHCISAVLLTDDGATIVTGCHDGOICLWDLSVELQINPRALL	60
Db	1	MAGNSLVLPVLWGKAPTHCISILLTDDGTTIVTGCHDGOICLWDLSVELQINPRALL	60
QY	61	FIGHTASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLACTHTGIGFYQFS	120
Db	61	FIGHTAITCLSKACASGDQYIVTSASANGEMCLWDVNDGRCIEFTKLACTHTGIGFYQFS	120
QY	121	VGNQREGRLCHGHYPEILVVDATSLVLYSVLSKISPDWISSMSIIRSHRTQEDTVVAL	180
Db	121	VGNQREGRLCHGHYPEILVVDATSLVLYSVLSKISPDWISSMSIIRSHRTQEDTVVAL	180
QY	181	SVTGILKWIIVTSEISDMQDTEPIPEEESKPIYCONCOSISFCATFQRSLLVVCSEYRV	240
Db	181	SVTGILKWIIVTSEISGQDTEPIPEEESKPIYCONCOSLSFCATFQRSLLVVCSEYRV	240
QY	241	FDAGYSLLCSGSPSNGQTWCGDFVSDKVIIWTEGQSYIYKLPASCLPASDSFRSDV	300
Db	241	FDAGYSLLCSGSPSDGQTWCGDFVSADKVIWTEGQSYIYKLPASCLPASDSFRSDV	300
QY	301	GRAVENLPPVQHILLDRKDELLICPPVTRFPYCREYFHKLLIQGSSGRLNTWNI	360
Db	301	GRAVENLPPVQHSLDDQKRELVI CPPVTRFPYCKEYLHKLLIQGSSGRLSIWNI	360
QY	361	TADKQGSBEGLAMTTSISLOEAPDKLPCPAGIIDQLSVIPNSNPLKVTASVYPAHGR	420
Db	361	IADKQANEGLKTTTCISLODAPDKLPCPAGIIDQLSVIPNSNPLKVTASVYPAHGR	420
QY	421	LVCGREDGSIIVPATQTAIVQLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY	480
Db	421	LVCGREDGSIIVPATQTAIVQLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY	480
QY	481	DQRYLISGGVDPSVIWIDIFSGEMKHI FCVHGGEITQLLVPPENCARSVOHCICSVASDH	540
Db	481	DQRYLISGGVDPSVIWIDIFSGEMKHI FCVHGGEITQLLVPPENCARSVOHCICSVASDH	540
QY	541	SVGLLSLREKKGICMLASRHLFPQVWKRPSPDYLIVVGCSDGVSVMQMDTGALDRCVMG	600
Db	541	SVGLLSLREKKGICMLASRHLFPQVWKRPSPDYLIVVGCSDGVSVMQMDTGALDRCAMG	600
QY	601	ITAVEILNACDEAPVPAVDSLHPAVNLQAMTRRSALAKKMAHKKLQTLATNLLASEA	660
Db	601	ITAVEILNACDEAPVPAVDSLHPAVNLQAMTRRSALAKKMAHKKLQTLATNLLASEA	660
QY	661	SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTPAESRPTALISPNL	720
Db	661	SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTPAESRPTALISPNL	720
QY	721	QKAGSSDKGSGFLTGKRAAVLFQOVKETIKENIKHEHLDDDEEIEIMRQREESDPY	780
Db	721	QKAGSSDKGSGFLTGKRAAVLFQOVKETIKENIKHEHLDDDEEIEIMRQREESDPY	780
QY	781	RSSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPCHTVSFGLLSR	840
Db	781	RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPCHTVSFGLLSR	840
QY	841	GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBGVKGTYGVSRAVTTQHLISISLANT	900
Db	841	GGHMSLMLPGYNQAAKGLHAKAEVGRKLPABEGVGKGYTVSRAVTTQHLISISLANT	900
QY	901	LMSMTNATFIGDHMKKGTPRPPRPTDLSKARGSPPTSSNIVQGIQKQVAAPVVSARS	960
Db	901	LMSMTNATFIGDHMKKGTPRPPRPTDLSKARDSPPASSNIVQGIQKQ-----	949
QY	961	ADHSGSDPPSAPALHTCFVLNKGWSQLAAHMCVMLPDLGLGDKFRPPLLEMLARRWQDRC	1020

```

Db 950 -----GHSQALAMHCVMPLDLLGLGKFRPPLEMLARRWQDR 987
Qy 1021 LEVREAAQALLAEIRIEQAGREKADAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1080
Db 988 LEVREAAQALLAEIRIEQAGREKETTIDTAPYLPQYMDHVISPGVTAEAMQTWAAAPDA 1047
Qy 1081 SGPEAKVQEEHDLVDDDDITTCGLSSVPQMKKISTSEERKQATAIVLGVIGAEFGAE 1140
Db 1048 SGPEAKVQEEHDLVDDDDITTCGLSSVPQMKKISTSEERKQATAIVLGVIGAEFGAE 1107
Qy 1141 IEPKLLTRPRSSQIPEGFGLTSGGNSYLARHTCKALTPLLLQPPSPKLPHPHSTIRRT 1200
Db 1108 IEPKLLTRPRSSQIPEGFGLTSGGNSYLARHTCKALTPLLLQPPSPKLPHPHSTIRRT 1167
Qy 1201 AIDLIGRFTWPEYMDVSAYMLGELCADAEKQLANITMGLPLSPAADSARSARHALS 1260
Db 1168 AIDLIGRFTWPEYMDVSAYMLGELCADAEKQLANITMGLPLSPAADSARSARHALS 1227
Qy 1261 LIATARPAPFTTTAKEVHRHTALANTQSQNNHTTTLARAKGEILRVIEILIEKPTPD 1320
Db 1228 LIATARPAPFTTTAKEVHRHTALANTQSQSIHTTTTLARAKGEILRVIEILIEKPTPD 1287
Qy 1321 VVDLLVEVMDIIMYCLSGSLVKKGLQECFPAICRFYMWVSYYERNHRIAVGARHGSVALY 1380
Db 1288 VVDLLVEVMDIIMYCLSGSLVKKGLQECFPAICRFYMWVSYYERNHRIAVGARHGSVALY 1347
Qy 1381 DIRTKKQTIHGKGPITAVAFAPDGRYLATYSNTDISHISFWQNTSILGSLGMLNSAPO 1440
Db 1348 DIRTKKQTIHGKGPITAVAFAPDGRYLATYSNTDISHISFWQNTSILGSLGMLNSAPO 1407
Qy 1441 LRCIKTYQVPPVPASPQSHNALKLARLITWTSNRNVILMAHDGKEHRFMV 1490
Db 1408 LRCIKTYQVPPVPASPQSHNALKLARLITWTSNRNVILMAHDGKEHRFMV 1457

RESULT 7
WDR7 HUMAN STANDARD; PRT; 1160 AA.
AC Q9V4EG;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE WD-repeat protein 7 (Fragment).
GN Name=WD7; Synonyms=KIAA0541;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-35(1998)
CC -1- SIMILARITY: Contains at least 6 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB011113; BAA25467.1; -.
DR PIR; T00272; T00272.
DR Genew; HGNC:13490; WDR7.
DR InterPro; IPR001680; WD40.
DR InterPro; IPR011046; WD40_like.

```

```

DR Pfam; PF00400; WD40; 4.
DR PRINTS; PRO0320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 3.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT NON TER 1 1
FT REPEAT 1 36 WD 1.
FT REPEAT 74 113 WD 2.
FT REPEAT 132 177 WD 3.
FT REPEAT 228 267 WD 4.
FT REPEAT 1021 1060 WD 5.
FT REPEAT 1062 1102 WD 6.
SQ SEQUENCE 1160 AA; 127452 MW; A378B4351832484D CRC64;

Query Match 77.58; Score 6023; DB 1; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 331 RFFYGCREFYFKLLIQDSSGRNLNWNISDTADKQSGSEGLAMTTSISLQEAFLKLP 390
Db 1 RFFYGCREFYFKLLIQDSSGRNLNWNISDTADKQSGSEGLAMTTSISLQEAFLKLP 60
Qy 391 AGIIDQLSVIPNSNEPLKVTASVYIPAHGRVLCREDGSIIVPATQTATVQLLQGEHML 450
Db 61 AGIIDQLSVIPNSNEPLKVTASVYIPAHGRVLCREDGSIIVPATQTATVQLLQGEHML 120
Qy 451 RRGWPPHRTLRGHRNKVTCLLYPHQVSARYDQRYLISGGVDFSVIWDIFSGEMKHIFCV 510
Db 121 RRGWPPHRTLRGHRNKVTCLLYPHQVSARYDQRYLISGGVDFSVIWDIFSGEMKHIFCV 180
Qy 511 HGGEITQLLVPPENCARSARVOHCICSVASDSVGLLSLRREKCKIMLSRHLFPPIQVIKWRP 570
Db 181 HGGEITQLLVPPENCARSARVOHCICSVASDSVGLLSLRREKCKIMLSRHLFPPIQVIKWRP 240
Qy 571 SDDYLVVGCSDGSVYVQMDTGALDRCMVGITAVEILNACDEAVPAADVSLSHPAVNLKQ 630
Db 241 SDDYLVVGCSDGSVYVQMDTGALDRCMVGITAVEILNACDEAVPAADVSLSHPAVNLKQ 300
Qy 631 AMTRRSALAKNMAHKKLOTLATNLLASEADKGNLPKYSHNSLMVQAIKTNLTDPDIHV 690
Db 301 AMTRRSALAKNMAHKKLOTLATNLLASEADKGNLPKYSHNSLMVQAIKTNLTDPDIHV 360
Qy 691 LFFDVEALIIQLLTERASRNTALISPENLQKASGSDKGSFLTGKRAAVLFOQVKETI 750
Db 361 LFFDVEALIIQLLTERASRNTALISPENLQKASGSDKGSFLTGKRAAVLFOQVKETI 420
Qy 751 KENIKEHLDDDEEDEEIMRQREESDPEYRSKSKPLTLLEYNLTMDTAKLFMSCLHAW 810
Db 421 KENIKEHLDDDEEDEEIMRQREESDPEYRSKSKPLTLLEYNLTMDTAKLFMSCLHAW 480
Qy 811 GLNEVLDEVCLDRGLMKLPHCTVSFGLSRGSHMSLMPCYNQPACKLSHGKTEVGRKLP 870
Db 481 GLNEVLDEVCLDRGLMKLPHCTVSFGLSRGSHMSLMPCYNQPACKLSHGKTEVGRKLP 540
Qy 871 ASSEGVGKTYGVSRVATTQHLLSIIILANTLMSMTNATFIDGMKKGTPRPPRSTPDLS 930
Db 541 ASSEGVGKTYGVSRVATTQHLLSIIILANTLMSMTNATFIDGMKKGTPRPPRSTPDLS 600
Qy 931 KARGSPPTSSNIYVQGIKQVAAPVWSARSADHSGSDPPSPALHTCFLVNEGWSQLAAM 990
Db 601 KARGSPPTSSNIYVQGIKQVAAPVWSARSADHSGSDPPSPALHTCFLVNEGWSQLAAM 660
Qy 991 HCWMLPDLGLDKFRPPLLEMLARRWQDRCLFVREAAQALLAEIRIEQAGREKADAW 1050
Db 661 HCWMLPDLGLDKFRPPLLEMLARRWQDRCLFVREAAQALLAEIRIEQAGREKADAW 720
Qy 1051 APYLPQYIDHVISPGVTSEAAQTITAPDASGPEAKVQEEHDLVDDDDITTCGLSSVPQM 1110
Db 721 APYLPQYIDHVISPGVTSEAAQTITAPDASGPEAKVQEEHDLVDDDDITTCGLSSVPQM 780
Qy 1111 KKISTSYEERRKQATAIVLGVIGAEFGAIBPPLKLTTPRSSSQIPGEGFLTSGGSNYS 1170

```

```
Db 781 KKTSTSEERKQATAVLVGVGAEGAEIEPPKLLTRSSSQIPEGFGLTSGGSNY 840
Qy 1171 LARHTKALFLLQPPSPKLPKPHSTIRTRTAIDLIGRGFTVWPDVSAVLMGLLELCA 1230
Db 841 LARHTKALFLLQPPSPKLPKPHSTIRTRTAIDLIGRGFTVWPDVSAVLMGLLELCA 900
Qy 1231 DAEKOLANITWGLPLSPAADSASARHALSLIATARPAPITTTIAKEVHRHTALAANTOS 1290
Db 901 DAEKOLANITWGLPLSPAADSASARHALSLIATARPAPITTTIAKEVHRHTALAANTOS 960
Qy 1291 QQNHHITTLARAKEILRVIEIIEKMPDVLVVEVMDIIMYCLEGSLVKKKGQECF 1350
Db 961 QQNHHITTLARAKEILRVIEIIEKMPDVLVVEVMDIIMYCLEGSLVKKKGQECF 1020
Qy 1351 PAICRFYMSYERNHRIAGARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLA 1410
Db 1021 PAICRFYMSYERNHRIAGARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLA 1080
Qy 1411 TYSNTDHSIFWQNTSLGSGIWMNSAPOLRCIKTYQVPPVQSPGSHNALKLARLIW 1470
Db 1081 TYSNTDHSIFWQNTSLGSGIWMNSAPOLRCIKTYQVPPVQSPGSHNALKLARLIW 1140
Qy 1471 TSNRNVLMAHDGKEHRFMV 1490
Db 1141 TSNRNVLMAHDGKEHRFMV 1160

RESULT 8
Q80TY3
ID Q80TY3 PRELIMINARY; PRT; 1166 AA.
AC Q80TY3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MKIAA0541 protein (Fragment).
DE Name=Wdr7; Synonyms=mkIAA0541;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT ii. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
CC -1- SIMILARITY: Contains 4 WD repeats.
DR EMBL; AK122304; BAC65586.1; -.
DR MGD; MGI:1860197; Wdr7.
DR InterPro; IPR001680; WD40.
DR IncePro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT NON TER
SQ SEQUENCE 1166 AA; 128360 MW; 0D4B66A78F262D4D CRC64;

Query Match 74.48; Score 5784.5; DB 2; Length 1166;
Best Local Similarity 92.98; Pred. No. 0;
Matches 1114; Conservative 27; Mismatches 33; Gaps 1;

Qy 292 ASDSFRSDVGKAVENLIPVQHSLLDQDKDELVICPPVTRFFYGCCKEYLHKLIIQDSSG 351
|||||
```

```
Db 1 ASDSFRSDVGKAVENLIPVQHSLLDQDKDELVICPPVTRFFYGCCKEYLHKLIIQDSSG 60
Qy 352 RLNIWNIISTADKQSGSEGLAMTTISLSQAFDKNPCAGIIDQLSVLPNSNEPKVTA 411
Db 61 RLNIWNIADIAEQEADKQSGSEGLAMTTISLSQAFDKNPCAGIIDQLSVLPNSNEPKVTA 120
Qy 412 SVYIPAGRLVCGREGDSIVIVPATQTAIVQLIQGHMLRRGPPHRTLRGHNKVTCLL 471
Db 121 SVYIPAGRLVCGREGDSIIIVPATQTAIVQLIQGHMLRRGPPHRTLRGHNKVTCLL 180
Qy 472 YPHQVSARYDQRYLISGGVDFSVIIMDFSGEMKHIFCVHGGEITOLLVPPENC SARVQH 531
Db 181 YPHQVSARYDQRYLISGGVDFSVIIMDFSGEMKHIFCVHGGEITOLLVPPENC SARVQH 240
Qy 532 CICSVASDHSVGLISREKKIMLASRHLPFOVIKWRPSDDYLVVGCSDGSGVYVQMDT 591
Db 241 CICSVASDHSVGLISREKKIMLASRHLPFOVIKWRPSDDYLVVGCSDGSGVYVQMDT 300
Qy 592 GALDRVCVMGITAVEIILNACDEAVPAAVDSLSHPAVNLKQAMTRRSIILAKNMAHKLQTL 651
Db 301 GALDRVCVMGITAVEIILNACDEAVPAAVDSLSHPAVNLKQAMTRRSIILAKNMAHKLQTL 360
Qy 652 ATNLLASEASDKGNLPKYSHNSLMVQAIKTNLTDPIHVLFPDVEALIIQLITEASRPN 711
Db 361 ATNLLASEASDKGNLPKYSHNSLMVQAIKTNLTDPIHVLFPDVEALIIQLITEASRPN 420
Qy 712 TALISPENIQKSGSGSDKGSFLTGKRAAVLFOOVKETIKENIKHLLDDEEDEEIMRQ 771
Db 421 TALISPENIQKSGSGSDKGSFLTGKRAAVLFOOVKETIKENIKHLLDDEEDEEIMRQ 480
Qy 772 RRESDPEYRSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDLRLGLMKPHC 831
Db 481 RRESDPEYRSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDLRLGLMKPHC 540
Qy 832 TVSFGLLSRGHNSLMPLPGYNQAPACKLSHGKTEVGRKLPASEGVGKGTGVSAVTTQHL 891
Db 541 TVSFGLLSRGHNSLMPLPGYNQAPACKLSHGKTEVGRKLPASEGVGKGTGVSAVTTQHL 600
Qy 892 LSTIISLANTLMSMTNATFIDGDMKGPTRPPRSTPDLKARGSPPTSSNIVOGQIKOVA 951
Db 601 LSTIISLANTLMSMTNATFIDGDMKGPTRPPRSTPDLKARGSPPTSSNIVOGQIKOVA 658
Qy 952 APVVSARSADHSGSDPPSAPALHTCFLVNEGWSQLAAMHCVMLPDLGLGLDKFRPPLLEM 1011
Db 659 -----GWSQLAAMHCVMLPDLGLGLDKFRPPLLEM 687
Qy 1012 LARRWQDRCLVEVREAAQALLLAEELRIEAGRKEATDAWAPYLPQIVDHSVPGVTSEAA 1071
Db 688 LARRWQDRCLVEVREAAQALLLAEELRIEAGRKEATDTWAPYLPQIVDHSVPGVTSEAA 747
Qy 1072 QTITTPADASGPEAKVQEEHDLVDDDIITTCGLSSVPQMKCISTSYEERRKQATAVLILG 1131
Db 748 QTMAAADASGPEAKVQEEHDLVDDDIITTCGLSSVPQMKCISTSYEERRKQATAVLILG 807
Qy 1132 VIGAEGFAGIEPPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALTALLQPPSPKL 1191
Db 808 VIGAEGFAGIEPPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALTALLQPPSPKL 867
Qy 1192 PPHSTIRTRTAIDLIGRGFTVWPDVSAVLMGLLELCADEKQALANIITWGLPLSPAADS 1251
Db 868 PPHSTIRTRTAIDLIGRGFTVWPDVSAVLMGLLELCADEKQALANIITWGLPLSPAADS 927
Qy 1252 ARSARHALSLIATARPAPITTTIAKEVHRHTALAANTOSQONMHTTTLARAKEILRVIE 1311
Db 928 ARSARHALSLIATARPAPITTTIAKEVHRHTALAANTOSQONMHTTTLARAKEILRVIE 987
Qy 1312 ILIEKMPDVLVVEVMDIIMYCLEGSLVKKKGQECFPAICRFYMSYERNHRIAVG 1371
Db 988 ILIEKMPDVLVVEVMDIIMYCLEGSLVKKKGQECFPAICRFYMSYERNHRIAVG 1047
Qy 1372 ARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLATYSNTDHSIFWQNTSLGSG 1431
Db 1048 ARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLATYSNTDHSIFWQNTSLGSG 1107
```

QY 1432 IGMNSAPQRCIKTYQVPPVQSPASPGSHALKLARLIWTSNRNVLMAHDGKEHRPMV 1490
DB 1108 IGMNSAPQRCIKTYQVPPVQSPASPGSHALKLARLIWTSNRNVLMAHDGKEHRPMV 1166

RESULT 9
Q6GPQ7 Q6GPQ7 PRELIMINARY; PRT; 956 AA.
AC Q6GPQ7;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE MGC82685 protein.
GN Name=MGC82685;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.K., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.B., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 5 WD repeats.
DR EMBL; BC073055; AAF73055.1; -.
DR InterPro; IPR01680; WD40.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPOTEINERPT.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 956 AA; 105601 MW; 23A89D893460F923 CRC64;

Query Match 55.7%; Score 4331; DB 2; Length 956;
Best Local Similarity 85.1%; Pred. No. 1.1e-280;

Matches 808; Conservative 73; Mismatches 68; Indels 0; Gaps 0;
QY 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVVTGCHDQICLMDLSVELQINPRALL 60
DB 1 MAGNSLVLPVLMGRKAPTHCISTVLVMDLLETITVGTCHDQICLMDLSPELEITPRALL 60
QY 61 FGTASITCLSKACASSDKOYIVSASESGEMCLMDYSDGRCIEFTKLACTHTTGIOYQFS 120
DB 61 FGTASITCLSKACASSDRQYIVSASESGEMCLMDVNDGRCVETFKLACTHTTGIOYQFS 120
QY 121 VGNQREGRLCHGHYPEILLVWDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
DB 121 IGTQREGRLCHGHYPEVLVMDASSLDVLYVSKISPDWISSMSIIRSHRTQEDTVAV 180
QY 181 SVTGILKWIWTVTSEISDMQDTEPIFEESKPIYQCNQCSISFCAPTORSLLVVCVKYRW 240
DB 181 SVTGILKWIWTVTAEVHRMQDTEPVPEESKPIYQCNQCSISFCAPTORSLLVVCVKYRW 240
QY 241 FDAGDYSLICSGSPSENGQTWTGDDFYSSDKVLIWTENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 LDAGDFSLICSPSEEGQTWTGDDFIADKVIATWEDGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIPPVOHILADKXKELLICPPVTRPFVGCVEYFHKLLIQDSSGRININWISD 360
DB 301 GRAIENSIPPLYSLYIMPDKDKLLICPPVTRPFYGRKSFHKLLIQDSSGRICMWSVPD 360
QY 361 TADKQSGSEGLAMTTSISLQEAPDKLNPAGIIDQLSVIPNSNEBLKVYVYIPAHGR 420
DB 361 ALEQLENKQGLQITTTTSLQEAPDKLNPAGIIDQLSIMPNSVEPLKVYVYIPSHGR 420
QY 421 LVCGREDGSIIVPATQTAIVQLQGEHMLRGWPPHRTLGRHNRKVTCLLYPHQVSARY 480
DB 421 LVCGREDGSIIVPATQTAIVQLQGEHMLRGWPPHRTLGRHNRKVTCLLYPHQVSARY 480
QY 481 DQRYLISGGVDSVIVWIDIFSGEMKHFVHGGEITQLLVPPENCASARVOHCICSVASDH 540
DB 481 DQRYLISGGVDSVIVWIDIFSGEMKHFVHGGEITQLLVPPENCASARVOHCICSVASDH 540
QY 541 SVGLLSLRKCKCIMLASRHLFPQVWKRPSPDDYLVVGCSDGVSYYVQMDTGALDRCVMG 600
DB 541 SVGLLSLRKCKCIMLASRHLFPQVWKRPSPDDYLVVGCSDGVSYYVQMDTGALDRCVMG 600
QY 601 ITAVEILNACDEAPVPAVDSLSHPAVNLQAMTRRSIALKQMAHKLQTLATNLASEA 660
DB 601 ITAVEILNACDEAPVPAVDSLSHPAVNLQAMTRRSIALKQMAHKLQTLATNLASEA 660
QY 661 SDKGNLPKYSRNSLMVQAIKTNLTDPDIHVLFDVEALIIQLLTERASRNTALISPENL 720
DB 661 SDKGNLPKYSRNSLMVQAIKTNLTDPDIHVLFDVEALIIQLLTERASRNTALISPENL 720
QY 721 QKASGSSDKGSGFLTGKRAAVLFQVVKETIKENIKEHLDDDEDEBEIMRQREESDPEY 780
DB 721 HKASGSSDKGSGFLTGKRAAVLFQVVKETIKENIKEHLDDDDDDDDAMRQKEDSDPEY 780
QY 781 RSSKSKPLTLLEYNLTMTAKLFMSCLHAWGLNEVLDEVLDRGLMKPHCTVSGFLLSR 840
DB 781 RTSKSKPLTLLEYNLTMTAKLFMSCLHAWGLNSALDDVCLERLGLMKPHCTVSGFLLSR 840
QY 841 GGHMSLMPLGYNQPACKLSHGKTEVGRKLPASEGVCKGTGYGSRATTOHLLSIISLANT 900
DB 841 GGHMSLMPLPRFNQTVYKASDGNVIGRKMSISEGLGKGTGYGSRATTOHLLSIISLANT 900
QY 901 LMSMTNATFIDHMKKGPTRPPRSPDLSKARGSPPTSSNIVQOQIKQ 949
DB 901 LMSMTNATFIDHMKKGPTRPPRSPDLSKARGSPPTSSNIVQOQIKQ 949

RESULT 10
Q86VP2 Q86VP2 PRELIMINARY; PRT; 614 AA.
AC Q86VP2;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

```
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE WDR7 protein (Fragment).
GN Names=WDR7;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RA Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 WD repeat.
DR EMBL: BC050352; AH50352.1; -.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR011046; WD40.
DR Pfam: PF00400; WD40; 1.
DR SMART: SM00320; WD40; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW WD repeat.
FT NON_TER
SQ SEQUENCE 614 AA; 67008 MW; 9A8E32222A6D15709 CRC64;

Query Match 40.8%; Score 3170; DB 2; Length 614;
Best Local Similarity 100.0%; Pred. No. 2.9e-203;
Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 877 KGTGYGSRVATTOHLLSIISLANTLMSMTNATFIGDHMKKGTPRPPRSTPDLKARGSP 936
DB 1 KGTGYGSRVATTOHLLSIISLANTLMSMTNATFIGDHMKKGTPRPPRSTPDLKARGSP 60

QY 937 PTSSNIVQGOIKQVAPVVSARGSDADHSQSDPPSPALHTCFLVNEGWSQLAAHWCMLP 996
DB 61 PTSSNIVQGOIKQVAPVVSARGSDADHSQSDPPSPALHTCFLVNEGWSQLAAHWCMLP 120

QY 997 DLGLGDKFRPPLLEMLARRWQDCLVREAAQALLAELRIEIQAGKEAIDAWAYLPQ 1056
DB 121 DLGLGDKFRPPLLEMLARRWQDCLVREAAQALLAELRIEIQAGKEAIDAWAYLPQ 180

QY 1057 YIDHVISPGVTSEAAQTITTPADSGPEAKVQEEHDLVDDDTTTCGLSSVPQMKKISTS 1116
DB 181 YIDHVISPGVTSEAAQTITTPADSGPEAKVQEEHDLVDDDTTTCGLSSVPQMKKISTS 240

QY 1117 YEERRKQATVILLVIGAFGEAIEPPKLLTTPRSSQITPEFGLTSGSGNSYLARHTC 1176
DB 241 YEERRKQATVILLVIGAFGEAIEPPKLLTTPRSSQITPEFGLTSGSGNSYLARHTC 300

QY 1177 KALTFTLLQPPSPKLPHPHSTIRRTAIDLIGRGFTVWPEYMDVSAVLMLGLELCADEAKQL 1236
```


Db 1032 IREAAQIILLGELTRMGKKRQKQVLSWAQYLPYLYTH--TEPIVGAQQQALALISQPASGG 1089
QY 1079 -----DASGPEAKV-----QBEHDLVDDDIITGCLSVPM 1110
Db 1090 AGSGGGGNGGVGVSGGGAGSGGPGGVPGGDAHQDEYEBEEI : : : : : 1139
QY 1111 KKTSTSYEERKQATAIVLLGVIGAEFGABE --BPPK-----LLTRPRS 1152
Db 1140 KPESSISELKKRKTAVILLGVIGAEFGQDISQESPNHRGSIEMATGANLTSVAGGER 1199
QY 1153 SSQIPGFGLTSGGSNSYLARHTCKALTFLLOPPSPKLPPLPHSTIRRTAIDLGRGFTVM 1212
Db 1200 KSSVVEGFGAN-----NLARLTSMALHLLYAPPSPKLPQYTPPLARAAIDLGRGFTVM 1254
QY 1213 EPMDVSVMGLLELCADAEKQLANITWGLPLSPAADSARSARHLSLIATARPFAFIT 1272
Db 1255 EPLYDVSKVLLGLLEISCEG-KAVPNLNYKPLTPQADACTARHALRLIATARPAFIT 1313
QY 1273 TIAKEVHRHTALAAQTQQNMHT-TTLARAKGEILRLVTEILIEKMPDQVVDLLVEVMDI 1331
Db 1314 TMAREVARVNTMOQAQSIINTPLTQSVLHKAKGEILQCVEMLIDRKQSEIAGLLVEVMDI 1373
QY 1332 IMYCLGSLVKKGLQECFPAICRFYMYVSYERNHRIAVGARHGSVALYDIRTKCQTIH 1391
Db 1374 ALHCVDGNELKNRGLAELCPAICKFNQISHCAQTRRIAVGANGNLAIYELRQNKQWIP 1433
QY 1392 GHGKPTAVAFAPDGRYLYATNSYNDSHISFWQNTSLLSGIMNSAPQLRCITQVPP 1451
Db 1434 AHTHPITSLAFSPDGKYLVSYSACENRLSPWQSTQMGF-LGQ----SOTRCTKGYSTAP 1488
QY 1452 VQPASPGSHNALKLRLIWTNSRNVLIMAHDGKEHRFMV 1490
Db 1489 IPDVS--RLNPMRLAKLWVNNRTVILMLADGSETRFNV 1525

RESULT 12

Q7Q3S0 PRELIMINARY; PRT; 1473 AA.
ID Q7Q3S0 AC Q7Q3S0; DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE EbiP7067 (fragment).
GN Name=ebic7067; ORFNames=ENSANGG00000005326;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 5 WD repeats.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008964; EAA12375.1; -.
DR InterPro; IPR001680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE; PSS0082; WD_REPEATS_2; 2.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT NON_TER 1473 1473
SQ SEQUENCE 1473 AA; 163970 MW; 5074D8D5F82A1AA7 CRC64;
Query Match 40.0%; Score 3107.5; DB 2; Length 1473;
Best Local Similarity 42.1%; Pred. No. 1.8e-198;
Matches 659; Conservative 285; Mismatches 452; Indels 171; Gaps 35;

```
Db 1012 MARWQHCLREIRNAOQLLGLGRMGKGRKQLVESWAOYLPMYTHTHPIVQOAPATG 1071
      :||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1068 SEAAQTITAPDAS-GPEAKVQEEHDLVDDITTCGLSSVPQMKKISTSYEBERRKOATA 1126
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1072 QSNAGSPSTSGTQPGLEHEESEBEV-----VRKPSSLAELKRRQTTA 1117
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1127 IVLGVIGAFGAIEPP--KLTRPRSSQIPGEGPLTSGGNSYSLARHTCKALTPLLL 1184
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1118 VLLGVIGAFGGQDISATDGKRSNENRRKSSVVEFGI-----GNNNLARSTSMALTHLL 1173
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1185 QPSPKLPPTSTTRTAIDILIGFTVWEPMVDVSAVLMGLLELCADABKOLANITWGLP 1244
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1174 APATQKLPATPURAADILIGFTVWEPIIDVSKVLLGLECCDSNKLPIESLVKLP 1233
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1245 LSPAADSARSARHLSIATARPAPFITTAKVHRHTALAANTQS--QQNMHTTTLARAK 1303
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1234 LTPQADACTARHALRIATARPAAPITTWAREVARVNTWQNAQAISVPIQTOSVLHRAK 1293
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1304 GEILRVIEILIEKPTVDVLLVEWMDIMYCLEGLSVKKGLQECPPALCRFPMVSYE 1363
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1294 REILQCVELMDIKMOTIEISNLVEWMDITLHCVDSDGLKKNGLAEVSPLMCKFNQVSHCS 1353
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1364 RNHRIAGARGHVALVDITGTCOTIHKHGPITAVAPADGRYLATYNTDHSIFWQ 1423
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1354 ASRIAGASGNHAIYELKQNKCOMIPATKQVATALAFSPDGKFLVSYCTENRSLFWQ 1413
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1424 MNTSLGSIGLNSAPQRCIKTYQVPPQVSPGSHNALKLARLWTSNRNVLMAHDG 1483
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1414 TSAGMFG-LGQ-----SQTRCIKGYSTAFIPDI--GRLNPMRLAKLIWINNRTVTLMLADG 1466
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1484 KEHRFMV 1490
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1467 SETRFNV 1473
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
```

RESULT 13

```
046092 PRELIMINARY; PRT; 1471 AA.
AC 046092;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE EG:86E4.3 protein.
GN Names:EG:86E4.3; ORFNames=CG17766;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferraz C., Vidal S., Brun C., Bucheton A., Demaille J.G.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 5 WD repeats.
DR EMBL; AL021086; CAAL5934.1; -.
DR FlyBase; FBgn0023510; CG17766.
DR InterPro; IPR011680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE; PS00682; WD_REPEATS_2; 1.
DR Repeat; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 1471 AA; 162496 MW; C468F629250DB2F2 CRC64;
```

Query Match

37.9%; Score 2944; DB 2; Length 1471;

```
Best Local Similarity 40.3%; Pred. No. 1.6e-187;
Matches 622; Conservative 294; Mismatches 447; Indels 182; Gaps 31;
Qy 54 INPRALLFGHTASITCLSKACASSDQYIVSASEGEMCLWDVSDGRCIEFTKLACTHTG 113
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 1 MSPRCLLVGHSAPVLCIVRASLLPENFLVSSSENGEMCTWDLTDGKCEAVKLPQVHTQ 60
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 114 IQYQSVGNQREGRLLCHGHYPEILLVVDATSLVLYSLVSKISPDWISMSIIRSHRTQ 173
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 61 IQSYH--TANSEDRVLCFYAEIIMVDPFSLVYVLSKVKPDPWISAIHRLPMRRK 118
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 174 EDTVVALSVTGILKMWITSEISDMODTRPIREESKPIYQNCQSIISFCAPTORSLLVV 233
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 119 DDVLAITTTGVVKWTLTG--NENKHAEPYENESKEIRCLNAINTWCAQNRQVTVLV 176
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 234 CSKYRWFVDAGDYSLLCSGSPENGQWTGDFVSSDKVIIWTENGOSIYIKLPAJCLPAS 293
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 177 CTYQWQIYDAGDFTVLCVAPARERWQSGDFITSRVMLWTDGKGLYKLPANCI 236
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 294 DSRSDVGKAVENLIPVQVHILLDRKXELLICPPVTRFPYGYGREFPHKLLIOGDSGRL 353
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 237 KEFHS---KSVVRDAPLYLVLOHAGDK-VLSCPPAMKLLIQAGAGQHN--LLRGDSGYI 290
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 354 NIW-----NLSDTADKQSGEGLAMTTSISLQEAFOKLNPCPAGIIDLVSIPNSNE 405
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 291 SVNWPEVPLDNIISIIQAKMPPLPKPHVCTSLVSAWSIMDPPPVGILQOLSRITES-- 348
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 406 PLKVTASVTIPAHGRVLCGREGDSIVIVPATQTAIVQLLQGEHMLRGGWPHRTLRHNRN 465
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 349 PVKLTSSYLPQOSRLVIGREGDSIVIVPATQVVMQLLVGIKQNFSDWFHQLYCHRG 408
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 466 KVTCLLYPHQVSAARYDQRYLISGVDPFVIIWIDIFSGEMKHIPCVHGGEITQLLVPPENC 525
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 409 RVNCLLCPSMIHSRYEKSHLLSGGIDFAVCLMDLYSGSLIHRFCVHAGEITQLLVPPESC 468
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 526 SARVQHCICSVASDHSVGLSLREKCIIMASHLEPIQVIKWRPDDVVLVWCSGDSVY 585
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 469 SPRILKICISVASDHSVTLVLSQERKCVTLASRHLFPVVTIKWAPLDDFLIVGCSGVS 528
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 586 VQOMDTGALDRCVMGITVAEILNACDE-----AVPAAVDSLSPHVNKLQA 631
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 529 VQOMETGHLDRVLHGLMAEVLVSACDEQAEDEGGSGGGSGNGASASEMGWNPVAFVFRG 588
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 632 MTRRSALAKNMAHKKLQTLATNLALASEADKGN---LPKYSHNSLMVQAIKTNLTDPDI 688
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 589 LKSRNNNAIR---HATQRTQIQLQLOQNGNFDPLMKHRSNPLVIQGLRTPNKDAES 644
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 689 HVLFPDVEALIIQLLTERASRNTALISPENLOKASGSDKSGSLTGTKEAAVLFOQVK- 747
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 645 HILFFDIEGLIFELHSEEYAOPTATLESGLVHLQNPDKGKMHLDASKKIGFFNKVKV 704
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 748 -----ETIKENIKHLLDDDEDEBIMR-----QRRESDPEYRSSSKPLTL 791
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 705 KAVDVEKILKDXKGLGVKQFKTEIIVEKVOAKVESLQKAVEPHEEQDLKASKTAKM 764
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 792 EYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRIGMLKPHCTVSGLLSRGGHSMMLPGY 851
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 765 EVTHVMEVAQLLSLLHSGLOPHLDKMCETRLGLLRPIVPIYSYGVLSKAGYNSLLPTW 824
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 852 NQ-----PACKL--SHGKTEVGRKLPASEGV-----GKGTYGVSRAVTTQHLISISLAN 899
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 825 QNNYAIPPGIQLPSSSKGKPLPEELORLEHTAVFTSRHLWELSTLTTHNLALVAMSN 884
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 900 TLMSWTNATFIGDHMKKGPTRPPRSTPDLKSARGSPPTSSNIYQQQIKQVAAAPVVSARS 959
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 885 TLLMSAASFL-----PDSEKHK-----KLQRLA-----QRT 911
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 960 DADHSGSDPPSAPALHTCFLVNEGNSQLAAMHCWMLPD--LLGLDKFPPLPLEMLARRW 1016
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Db 912 DSTLSNEEBREELMAHHSIQIKHAWSLTATHCFLPLDPKIEALEPKPKRQVEMVWGV 971
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
Qy 1017 QRCLEVRQAQALLLAELRRIEQAGKEAIDAWAPYLQYIDHVLISPGVTAAQTITT 1076
      :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
```

Db 972 QHHCIEIRAAQQILLGELTRMGKGRKQLVESAQVLYPLTH--TEPIVGAQQQLALIS 1029
Qy 1077 AP-----DASGPEAKV-----QSEHDLVDDDDITTGCL 1104
Db 1030 QPASGGAGSGGNGGVGVGSGGGAGSGGSGVPGGDAHQDEYEEEEEI----- 1084
Qy 1105 SSVPMKKISTSVERRKQATATVLLGVIGABFGAEI--RPPK-----L 1146
Db 1085 -----IRKPSLSSELKRKQTAVILLGVIGABFGQDISPNSHRSGISWATGANUTSGV 1139
Qy 1147 LTRPRSSQIPEGFGLTSGGSNSYLARHTCKALTFLLLOPPSPKLPHPHSTIRRTAIDLIG 1206
Db 1140 AGGERKSSVVEGFGIAN-----NLARLTSMALHLLYAPPSPKLPQYTPLEAAIDLIG 1194
Qy 1207 RGTWVEPMYMDVSAVLMGLELCADEKQOLANTMGLPLSPAADARSARSHALSILIATAR 1266
Db 1195 RGFTWVEPYLDVSKVLLGLEISECG-KAVPNLNLYKLPLTPQADACTARHARHALIATAR 1253
Qy 1267 PPAFITTIKEVHRHTALAAANTOSQONMHT-TTLARAKGEILLAVIELIEKMPDVTVDLL 1325
Db 1254 PAAFITTMAREVARYNTMQNAQSINTPLTQSVLHKAKGEILQCVEMLDKMQSEIAGLL 1313
Qy 1326 VEVMDIIMYCLEGLVKKGLQBCFPALCRFYMVSVYERNHRIAVGARHGSVALYDIRTG 1385
Db 1314 VEVMDIALHCVDCNGLKRLAEELCPALCKFNQISHCAQTRTAVGANGSLAIYELRQN 1373
Qy 1386 KCTIINGHKGPIITAVAFAPDGRYLAATYNTDISHISFWQMTSLGIGMNLNSAPQURCIK 1445
Db 1374 KCOMIPAHPTITSLAFSPDGRKYLVSACENRSLFWQSTGTFG-LGQ-----SOTRCK 1428
Qy 1446 TYQVPPVQVAPSGSHNALKLARLIWTSNRNVLMAHDGKEHRFMV 1490
Db 1429 GYSTAPIDVS--RLNPMRLAKLVNNRTVTLMLADGSETRFNV 1471

RESULT 14
Q8C711 PRELIMINARY; PRT; 531 AA.
AC Q8C711;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched
DE library, clone.D630037A10 product:hypothetical G-protein beta WD-40
DE repeats containing protein, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA The FANTOM Consortium.
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Haehizume W.,
RA Hayashida K., Hayatsu M., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tomaru A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: Contains 3 WD repeats.
CC EMBL: AK052730; BAC35119.1; --
DR InterPro; IPR001680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 3.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE; PS00682; WD_REPEATS_2; 2.
DR PROSITE; PS00294; WD_REPEAT_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 531 AA; 58749 MW; 8D9C5A61E66F0070 CRC64;
Query Match 34.3%; Score 2667; DB 2; Length 531;
Best Local Similarity 93.9%; Pred. No. 1.1e-169;
Matches 494; Conservative 19; Mismatches 13; Indels 0; Gaps 0;
Qy 1 MAGNSLVLPFLVLMGRKAPTHCISAVLLTDDGATIVTGCCHDQICLWDLSELQINPRALL 60
Db 1 MAGNSLVLPFLVLMGRKAPTHCISAVLLTDDGATIVTGCCHDQICLWDLSELQINPRALL 60
Qy 61 FGHASITCLSKACASSDKQYIVSASESGEMCLWVDSDGRCIEFTYKLAHTGTGTYQFS 120
Db 61 FGHASITCLSKACASGDKRYTVSASANGEMCLWVDNDGRCIEFTYKLAHTGTGTYQFS 120
Qy 121 VGNQREGRLCHGHYPEILLVDATSLVLYSVLSKISPDWISSMSTIRSHRTQEDTVVAL 180
Db 121 VGNQREGRLCHGHYPEILLVDATSLVLYSVLSKISPDWISSMSTIRSHRTQEDTVVAL 180
Qy 181 SVTGILKWIVTSEISDMQDTEPIFEESKPIYCNQCSISFCAPTRQLLVCSKYWRV 240
Db 181 SVTGILKWIVTSEMSGMQDTEPIFEESKPIYCNQCSISFCAPTRQLLVCSKYWRV 240
Qy 241 FDAGDYSLLCSPSENGQWTWGTGDFVSSDKVIITWENGOSYIYKLPASCLPASDSFRSDV 300
Db 241 FDAGDYSLLCSPSENGQWTWGTGDFVSSDKVIITWENGOSYIYKLPASCLPASDSFRSDV 300

QY 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFVCGCEYFHKLLIQDSSGRINMISD 360
DB 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFVCGCEYFHKLLIQDSSGRINMISD 360
QY 361 TADKQSEEGLAMTTSISLOAEFDKLNPCPAGIIDLQSVIPNSNEPLKVTASVYIPAHGR 420
DB 361 IAEKQEADEGLKMTTCISLOAEFDKLNPCPAGIIDLQSVIPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCGREGDSIIVPATQTAIVOLLQGHMLRRGWPPHRTLRGHRNKVTCILLYPHQVSARY 480
DB 421 LVCGREGDSIIVPATQTAIVOLLQGHMLRRGWPPHRTLRGHRNKVTCILLYPHQVSARY 480
QY 481 DQRYLISGGVDFSVIWIWDFSGEMKHIFCVHGGEITOLLVPPENC 526
DB 481 DQRYLISGGVDFSVIWIWDFSGEMKHIFCVHGGEITOLLVPPENC 526

RESULT 15
Q9BL05
ID Q9BL05 PRELIMINARY; PRT; 1335 AA.
AC Q9BL05;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Rab connectin related protein 2.
GN Name=rcb-2; ORFNames=Y54F10AM.10;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Ryan E., Courtney L., Yoakum M.;
RT "The sequence of C. elegans cosmid Y54F10AM.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;

Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
[9]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RN Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RN Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 6 WD repeats.
DR EMBL: AC025723; AAK29933.2; -.
DR WormBase; WBGene00004314; rbc-2.
DR WormPep; Y54F10AM.10; CE35950.
DR InterPro; IPR001680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SSQUENCE 1335 AA; 147103 MW; 19A5418BE1B65E07 CRC64;
Query Match 25.8%; Score 2006.5; DB 2; Length 1335;
Best Local Similarity 31.7%; Pred. No. 8e-125;
Matches 494; Conservative 264; Mismatches 491; Indels 307; Gaps 39;
QY 5 SLVPIVLWGRKAPTHCISAVLLTDDGATVTCCHDQICLWLSVLEQLNPRALLFGHT 64
DB 13 ALSVPVWVGPPENKICAIRVLPDGAIIITGAENGHIINWKLAEGLM--PKQLMIGH 70
QY 65 ASITCLSKACASDQKQYIVSASBSGEMCLWDSVGRGCIETFKLACTHTGTGTYQF-SVGN 123
DB 71 QKITAISQTTNTPTNTRFVSASADGRVCLWEIQDGRCIDSTSSIAHRIYIPYTKSRH 130
QY 124 QREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISMSIIRSHRTQEDTVVALSVT 183
DB 131 TRATRLFCIGDYSDIQWMDPQDITVYFSSISRRVPEPDWISCFITII-SHPEKEDQLIGMTLS 189
QY 184 GILKWIIVTSEISDMQDTEPIPEESKPIYCONCOSISFCAFTORSILVVCYKRVFDA 243
DB 190 GMMKWTLT-ELEKDPATSLYEDESKRLIEIQRISVSVCSTRRMILLIATCWMIDM 248
QY 244 GDYSLLCGPFSE-NGQTWTGGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDVGK 302
DB 249 DDLSTVVFHKNIEITTKRCVAGYLTDLKVTIGYDSEIHVFQLPISLQ-----KEGVPP 303
QY 303 AVENLIPVQHILLDRKDKELLICPPVTRFFVCGCEYFHKLLIQDSSGRINMISD 362
DB 304 VVANNRPTSNY--FDQPN-----PPIVATVNGVQHSRSL--NDVQFAFLPYNSNSL 353
QY 363 DKQSGEGLAMTTSISLOAEFDKLNPCPAGIIDLQSVIPNSNEPLK-----VTASVYIP 416
DB 354 ESSSS-----SYEK-----RRFSVVRSSRRDGGSVLNSDAVCCSLFVT 390
QY 417 AHGRVCGREDGSIIVPATQTAIVOLLQGHMLRRGW-----PPHRTLGRHKNKVT 469
DB 391 SQKFLGRSDGVIMTYACET-----LAQWLVRPAERATSRILTGHDAVRS 439
QY 470 LLYPHQVSARYDQRYLISGGVDFSVIWIWDFSGEMKHIFCVHGGEITOLLVPPENC 529
DB 440 MFYPFEHDTYDPQYFVSGDDFSDVIVMINSSTRITRFTVHGGPVKSFMLPPSNC 499
QY 530 QHCICSVASDHSVGLSLREKKICMLASRHLPIQVTKRPSDDYLVCSCDGSVVWQM 589
DB 500 TKCIASLAADNTIALNLIRDSKMLASRHPPIIQVWRPLDDFMLVKLADGSVVWQM 559

Qy	590	DTGALDRCVWGITAVEILNACDE--AVPAAVDSLS--HPAVNLKOAMTRRSIALAKN-----642
Db	560	ETANLDRIATGLLAEDITWATCDEQIGVEEGTDETSAHVAQLIRALKNKNEAVKQKVVG619
Qy	643	-----MAHHKLOT-LATNLLASEADKGNLPKYSHNSLMVQAIKTNLTDPDINHVP692
Db	620	GSVSGAATPSSHHELPTNCGTAVQLGSPMSILPLPGCAQGAHLVQ-----664
Qy	693	FDVEALLIQLITEASRPNTALISPE--NLQK-----ASGSSDKGGSFLTGTGRAAVLFQ744
Db	665	FEISALLTAGILHLDDSSAB-----APEKTTAQLDAAAAGASDSA---TGLSRKLTWQ715
Qy	745	QVKETIKENIKHLLDDDEEBEIMRORREESDEPYRSKXPIJTLLEYNLTMDTAKLFM804
Db	716	-----PEANLYLDVARLML729
Qy	805	SCLHAWGLNEVLDEVCLDCLMKLPHCTVSGLLSRGHHMSLMP-----GYNQPACKLSH860
Db	730	SMLHAWCLDEMDVCDKRLSHRPHQVYFGNVSRQELSVSLPTRFAADFESFCK---786
Qy	861	GKTEVGRKLPASEGVGKGTGVSRAVTOHLLSIIISLANTMSMTNATFIGDHMKGPTR920
Db	787	-----KSRWQASHSLNTSHLLAVIATSNITLMAKNSAL-----819
Qy	921	PPRESTPDLKARGSPPTSSNIVOGQIKQVAAPVVSARSADHSGSDPPSAPALHTCFLV980
Db	820	-----OLAKIRKSVEQNQSTIHQ-----PDRQQ-----I844
Qy	981	NECHSOLAHCVMCLPD-LLGLDKFRPLLEMLARRMOWDCLREVAQAALLAELERRIE1039
Db	845	KQGSVLAALHCVLLPDHPVPRSSYAPPSIEMLVRRWQDCLREIAQAALLIRETLRIG904
Qy	1040	QAQR-----KEADIMAPYLPQVIDHVISIPGVTSSAAQITITAPDASGPEAK1086
Db	905	AEGRFPSPDFFLKTVTRRLIESWTFPLPLDSDLSIFGSKLQSSVPTVQPSAPAP---961
Qy	1087	VQEEHDLVDDIITGCLSSVPQMKKITS-----VEERKQATVILGVIG1134
Db	962	-----PIPRTKNAPPDVEPVRGSEPTAEAGEAGIQOVRNQAOTSILLGVIG1010
Qy	1135	AEFGAEIEPPKLLTRPRSSQIPGFGLTSGGSNYSLARHTCKATLPLLQPPSPKLPHP1194
Db	1011	SQFGEEL-----NRADLTRATAVSELLELVAAENSLIPVH1045
Qy	1195	STIRRTAIDTIGRFTWPEYMDVSAVLMGLLELCADAЕК--QLANITMGLPLSPAADSA1252
Db	1046	SPLRRAIDLLGRGFVHWEHLEISKVVLGDLASNKPGTQRRIT-CAPLNAIEDAA1104
Qy	1253	RSARHLSLIATARPPAPITIAKEVHRHTALAANTOSQOMHTTTLARAKGEILLRVEI1312
Db	1105	RTSRQALSILALARPPLAITSLSMEVARYNA-AAQHOTIQHTVVSPLKRSSEVLRIIEE1163
Qy	1313	LIERKPTDVVDLLVEVMDIIMYCLEGSIVKKGIOEQCFPAICRFYMWYSYERHNRHTAVGA1372
Db	1164	LCEKRYNDIITAMLLPVGDILVHCLDITILKHKSMSVEFPPIPVKFNWVAYCSTKRVAFGG1223
Qy	1373	RHGSVALYDRTGKQCIHGHKGKGPITAVAPADGRYLATYNTDTSHTSFQWMTSLIGSI1432
Db	1224	KNGTCVVELRATKTHSLPSHNGPIAAVASEDGKYLATYGAEDGKINFQTQSFLGM-1282
Qy	1433	GMLNSAPOLRCIKTYQVPPVQVQSPSGSHNALKIARLIWTNRRNVILMAHDKGKHPR1488
Db	1283	-----GQOLKLTQSOPAPTYVSVPTTSGTSGFR-PRLVWINSKLSITMLMPGREGQRF1333

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:16:31 ; Search time 195 Seconds
(without alignments)
2947.872 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773

Sequence: 1 MAGNSVLPIVLWGRKAPTH.....TSNRNVILMAHDGKHEFRMV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7773	100.0	1490	17	US-10-645-335-2
2	7763	99.9	1490	15	US-10-093-463-154
3	2805.5	36.1	559	15	US-10-094-749-2856
4	1442	18.6	775	15	US-10-108-260A-3494
5	412.5	5.3	1391	16	US-10-437-963-162857
6	228.5	2.9	579	15	US-10-369-493-18893
7	228.5	2.9	608	15	US-10-369-493-20224
8	228	2.9	703	15	US-10-425-114-58817
9	228	2.9	710	16	US-10-425-115-360670
10	220.5	2.8	1005	15	US-10-369-493-20242
11	205.5	2.6	626	15	US-10-369-493-20247
					Sequence 2, Appli
					Sequence 154, App
					Sequence 2956, Ap
					Sequence 3494, Ap
					Sequence 162857,
					Sequence 18893, A
					Sequence 20224, A
					Sequence 58817, A
					Sequence 360670,
					Sequence 20242, A
					Sequence 20247, A

12	204	2.6	891	16	US-10-437-963-153705	Sequence 153705,
13	202	2.6	1136	15	US-10-369-493-19046	Sequence 19046, A
14	197.5	2.5	429	15	US-10-369-493-11348	Sequence 11348, A
15	191.5	2.5	478	15	US-10-369-493-18905	Sequence 18905, A
16	189	2.4	494	16	US-10-478-197-35	Sequence 35, Appli
17	186	2.4	573	9	US-09-925-299-1015	Sequence 1015, Ap
18	186	2.4	573	10	US-09-925-299-1015	Sequence 1015, Ap
19	186	2.4	1356	13	US-10-077-111-10	Sequence 10, Appli
20	185.5	2.4	679	15	US-10-369-493-20059	Sequence 20059, A
21	184.5	2.4	514	11	US-09-987-701-2	Sequence 2, Appli
22	184.5	2.4	514	11	US-09-987-701-4	Sequence 4, Appli
23	184	2.4	721	17	US-10-732-923-7436	Sequence 7436, Ap
24	181.5	2.3	514	11	US-09-987-701-12	Sequence 12, Appli
25	178	2.3	610	15	US-10-369-493-20185	Sequence 20185, A
26	175	2.3	1073	17	US-10-732-923-7538	Sequence 7538, A
27	174.5	2.2	465	16	US-10-687-732-10	Sequence 10, Appli
28	174.5	2.2	579	15	US-10-369-493-6573	Sequence 6573, Ap
29	174.5	2.2	587	16	US-10-687-732-25	Sequence 25, Appli
30	174	2.2	478	16	US-10-425-115-240549	Sequence 240549,
31	171	2.2	395	16	US-10-363-829-446	Sequence 446, App
32	171	2.2	981	14	US-10-032-585-7401	Sequence 7401, Ap
33	169.5	2.2	358	17	US-10-732-923-7507	Sequence 7507, Ap
34	169.5	2.2	1140	15	US-10-369-493-19053	Sequence 19053, A
35	169.5	2.2	1246	16	US-10-437-963-169035	Sequence 169035,
36	167.5	2.2	485	14	US-10-132-744A-6	Sequence 6, Appli
37	165.5	2.1	1508	16	US-10-483-506-9	Sequence 9, Appli
38	164.5	2.1	419	15	US-10-425-114-61169	Sequence 61169, A
39	163	2.1	411	16	US-10-437-963-105334	Sequence 105334,
40	163	2.1	1194	9	US-09-876-667-2	Sequence 2, Appli
41	163	2.1	1194	14	US-10-141-618-10	Sequence 10, Appli
42	163	2.1	1205	9	US-09-876-667-16	Sequence 16, Appli
43	162.5	2.1	475	16	US-10-425-115-362607	Sequence 362607,
44	162.5	2.1	557	15	US-10-425-114-72721	Sequence 72721, A
45	162.5	2.1	958	14	US-10-128-714-8398	Sequence 8398, Ap

ALIGNMENTS

RESULT 1
US-10-645-335-2
; Sequence 2, Application US/10645335
; Publication No. US2005006430A1
; GENERAL INFORMATION:
; APPLICANT: Masakazu Takeuchi
; APPLICANT: Yoshimi Takai
; TITLE OF INVENTION: Rabconnectin-3-binding protein
; FILE REFERENCE: 2003946-0055
; CURRENT APPLICATION NUMBER: US/10/645,335
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: JP 2002-319521
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 2
; LENGTH: 1490
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-645-335-2

Query Match 100.0%; Score 7773; DB 17; Length 1490;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAGNSVLPIVLWGRKAPTHCISAVLLTDDGATVTCGHDGOICLWDLVSVEIQINPRALL	60
DB	1	MAGNSVLPIVLWGRKAPTHCISAVLLTDDGATVTCGHDGOICLWDLVSVEIQINPRALL	60
QY	61	FQHTASITCLSKACASSDKQIVVSASEGEMCLWDSGRCIEFTKLACTHTGICQFYQFS	120
DB	61	FQHTASITCLSKACASSDKQIVVSASEGEMCLWDSGRCIEFTKLACTHTGICQFYQFS	120
QY	121	VGNQREGRLCHGHYPILVDVATSLVLYSVLSKISPDWISSMIIIRSHRTOEDTVVAL	180
DB	121	VGNQREGRLCHGHYPILVDVATSLVLYSVLSKISPDWISSMIIIRSHRTOEDTVVAL	180

Db 121 VGNQREGRLLCHGHYPELLVVDATSLVLSVSKISPDWISSMSIIRSHRTQEDTVVAL 180
Qy 181 SVTGILKXWIVTSEISDMQDTEPIFEESKPIYVCONCQISFCAPTORSLVLLVCSKYWRV 240
Db 181 SVTGILKXWIVTSEISDMQDTEPIFEESKPIYVCONCQISFCAPTORSLVLLVCSKYWRV 240
Qy 241 FDAGDYSLLCSGSENGQWTGDFVSSDKVLIWTENGQSYIYKLPASCLPASDPSRSDV 300
Db 241 FDAGDYSLLCSGSENGQWTGDFVSSDKVLIWTENGQSYIYKLPASCLPASDPSRSDV 300
Qy 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFYGCREFYFKLLIQDSSSGLNIWNISD 360
Db 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFYGCREFYFKLLIQDSSSGLNIWNISD 360
Qy 361 TADKQSEGLAMTTSISLOEAFDKNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGR 420
Db 361 TADKQSEGLAMTTSISLOEAFDKNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGR 420
Qy 421 LVCGREDGSIIVPATQTAIVQLLQGEHMLRRCWPPHRTLRGHRNKVTCLLYPHOVSARY 480
Db 421 LVCGREDGSIIVPATQTAIVQLLQGEHMLRRCWPPHRTLRGHRNKVTCLLYPHOVSARY 480
Qy 481 DQYLSISGGVDFSVIIVDFISGEMKHI FCVHGGEITQLLVPPENC SARVQHICISVASDH 540
Db 481 DQYLSISGGVDFSVIIVDFISGEMKHI FCVHGGEITQLLVPPENC SARVQHICISVASDH 540
Qy 541 SVGLLSUREKCCIMLASRHLFPQIVIKWRPSDYL VVVGSDG SVVYVQMDTGALDCVMG 600
Db 541 SVGLLSUREKCCIMLASRHLFPQIVIKWRPSDYL VVVGSDG SVVYVQMDTGALDCVMG 600
Qy 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQAMTRRSIALKNMAHKLQTLATNLNLAESA 660
Db 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQAMTRRSIALKNMAHKLQTLATNLNLAESA 660
Qy 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLITEASRNPNTALISPENL 720
Db 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLITEASRNPNTALISPENL 720
Qy 721 QKASGSDKGGSLTQGRAAVLFOQVKETIKENIKHLLDDEBDEBEIMRQREESDPEY 780
Db 721 QKASGSDKGGSLTQGRAAVLFOQVKETIKENIKHLLDDEBDEBEIMRQREESDPEY 780
Qy 781 RSSKSPKLTILEYNTLMDTAKLFMSCLHAWGLNEVLDCDLRLGMLKPHCTVSGFLSR 840
Db 781 RSSKSPKLTILEYNTLMDTAKLFMSCLHAWGLNEVLDCDLRLGMLKPHCTVSGFLSR 840
Qy 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBGVGKGTGVGSRAVTTQHLLSIISLANT 900
Db 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBGVGKGTGVGSRAVTTQHLLSIISLANT 900
Qy 901 LMSMTNATFIGDHMKKGTPRPPRSTPDLISKARGSPPTSSNIVQGOIKQVAAPVVSARS 960
Db 901 LMSMTNATFIGDHMKKGTPRPPRSTPDLISKARGSPPTSSNIVQGOIKQVAAPVVSARS 960
Qy 961 ADHSGSDPPSAPALHTCFLNVEGWSOLAAMHCVMPLDLLGLDKFRPPLLEMLARRWQDR 1020
Db 961 ADHSGSDPPSAPALHTCFLNVEGWSOLAAMHCVMPLDLLGLDKFRPPLLEMLARRWQDR 1020
Qy 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPYLPQVIDHVISPGVTSEAAQTITAPDA 1080
Db 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPYLPQVIDHVISPGVTSEAAQTITAPDA 1080
Qy 1081 SGPEAKVQBEHDLVDDDIITGCLSSVPQMKKISTSYEERKQATAILLVIGVIGAEFGAE 1140
Db 1081 SGPEAKVQBEHDLVDDDIITGCLSSVPQMKKISTSYEERKQATAILLVIGVIGAEFGAE 1140
Qy 1141 IEPKLLTRPRSSQIPEGGLTSGGSNYSLARHTCKALTFLQLQPPSKLPHSTIRRT 1200
Db 1141 IEPKLLTRPRSSQIPEGGLTSGGSNYSLARHTCKALTFLQLQPPSKLPHSTIRRT 1200
Qy 1201 AIDLIGRGFTWPEYMDVSAVLMGLLELCADAEKQLANITMGLPLSPAADSARSARHALS 1260
Db 1201 AIDLIGRGFTWPEYMDVSAVLMGLLELCADAEKQLANITMGLPLSPAADSARSARHALS 1260

Qy 1261 LIATARPPAFITTTIAKEVHRHTALAANTOSQONMHTTTTLARAKGEILRVIEILIEKMPD 1320
Db 1261 LIATARPPAFITTTIAKEVHRHTALAANTOSQONMHTTTTLARAKGEILRVIEILIEKMPD 1320
Qy 1321 VVDLLVEVMDIIMYCLEGSLVKKKGLQECFPAICRFYVYVYERNHRIAVGARHGSVALY 1380
Db 1321 VVDLLVEVMDIIMYCLEGSLVKKKGLQECFPAICRFYVYVYERNHRIAVGARHGSVALY 1380
Qy 1381 DIRTGKCQTIHGHKGPIITAVAPADGRYLATVNTDSDHISFWQMTSLLSIGMGLNSAPQ 1440
Db 1381 DIRTGKCQTIHGHKGPIITAVAPADGRYLATVNTDSDHISFWQMTSLLSIGMGLNSAPQ 1440
Qy 1441 LRCIKITYQVPPVQSPASPGSHNALKARLIWTSNRNVILMAHDGKHEFRMV 1490
Db 1441 LRCIKITYQVPPVQSPASPGSHNALKARLIWTSNRNVILMAHDGKHEFRMV 1490

RESULT 2

US-10-093-463-154

; Sequence 154, Application US/10093463

; Publication No. US20030208039A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Shenoy, Suresh

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Gusev, Vladimir

; APPLICANT: Pochart, Pascal

; APPLICANT: Zhong, Mei

; APPLICANT: Rastelli, Luca

; APPLICANT: Mezes, Peter

; APPLICANT: Smithson, Glennda

; APPLICANT: Guo, Xiaojia

; APPLICANT: Gerlach, Valerie

; APPLICANT: Casman, Stacie

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Tchernev, Velizar

; APPLICANT: Gangolli, Esha

; APPLICANT: Vernet, Corine

; APPLICANT: Pena, Carol

; APPLICANT: Burgess, Catherine

; APPLICANT: Liu, Xiaohong

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gorman, Linda

; APPLICANT: Spaderna, Steven

; APPLICANT: Voss, Edward

; APPLICANT: Malyankar, Uriel

; APPLICANT: Anderson, David

; APPLICANT: Patturajan, Meera

; APPLICANT: Miller, Charles

; APPLICANT: Taupier, Raymond J. Jr.

; TITLE OF INVENTION: No. US20030208039A1e1 Antibodies that Bind to Antigenic Polypepti:

; FILE REFERENCE: 21402-290A (Cura 590AT)

; CURRENT APPLICATION NUMBER: US/10/093,463

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 60/283,675

; PRIOR FILING DATE: 2001-04-14

; PRIOR APPLICATION NUMBER: 60/338,092

; PRIOR FILING DATE: 2001-12-03

; PRIOR APPLICATION NUMBER: 60/274,281

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/274,101

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/325,681

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/304,354

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/279,995

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 60/294,899

```
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 154
; LENGTH: 1490
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-093-463-154

Query Match      99.98; Score 7763; DB 15; Length 1490;
Best Local Similarity 99.94; Pred. No. 0;
Matches 1488; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1  MAGNSLVLPVLMGRKAPTHCISAVLITDDGATIVTGDGQICLWDLSELQINPRALL 60
DB      1  MAGNSLVLPVLMGRKAPTHCISAVLITDDGATIVTGDGQICLWDLSELQINPRALL 60

QY      61  FGHASITCLSKACASSDKQYIVSASESGEMCLWDVSDGRGCIETFTKLACTHTGIGYQFS 120
DB      61  FGHASITCLSKACASSDKQYIVSASESGEMCLWDVSDGRGCIETFTKLACTHTGIGYQFS 120

QY      121  VGNOREGRLLCHGHYPEILVVDATSLVLYSLYSKISPDWISSMSIIRSHRTQEDTVALL 180
DB      121  VGNOREGRLLCHGHYPEILVVDATSLVLYSLYSKISPDWISSMSIIRSHRTQEDTVALL 180

QY      181  SVTGILKVMITVTEISDMQDTEPIFEESKPIYCONCQSIISFCAPTORSLVVCYSKYRV 240
DB      181  SVTGILKVMITVTEISDMQDTEPIFEESKPIYCONCQSIISFCAPTORSLVVCYSKYRV 240

QY      241  FDAGDYSLLCSGSPENGQWTGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300
DB      241  FDAGDYSLLCSGSPENGQWTGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300

QY      301  GKAVENLIPVQHLLDRKDKELLICPPVTRFPYGCREFPHKLLIQDSSGRNLINWISD 360
DB      301  GKAVENLIPVQHLLDRKDKELLICPPVTRFPYGCREFPHKLLIQDSSGRNLINWISD 360

QY      361  TADQSGSEGLAMTTSISLQAEFDKLNPCPAGIIDLVSIPNSNEPIKVTVASVYIPAGR 420
DB      361  TADQSGSEGLAMTTSISLQAEFDKLNPCPAGIIDLVSIPNSNEPIKVTVASVYIPAGR 420

QY      421  LVCGREDGSIIVPQTQTAIVQLLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY 480
DB      421  LVCGREDGSIIVPQTQTAIVQLLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY 480

QY      481  DQYILSGGVDFSVIIVDIPSGEMKHIFCVHGGEITQLLVPPENC SARVOHCICSVASDH 540
DB      481  DQYILSGGVDFSVIIVDIPSGEMKHIFCVHGGEITQLLVPPENC SARVOHCICSVASDH 540

QY      541  SVGLLSIREKKCTMLASRLHFPQVLKWRPSDDYLVVGCSDGVSYYQMDTGALDRCVMG 600
DB      541  SVGLLSIREKKCTMLASRLHFPQVLKWRPSDDYLVVGCSDGVSYYQMDTGALDRCVMG 600
```

```
QY      601  ITAVEITLNACDEAVPAAVDSLSHPAVNLKQAMTRSLAALKNMAHKLQTLATNLASEA 660
DB      601  ITAVEITLNACDEAVPAAVDSLSHPAVNLKQAMTRSLAALKNMAHKLQTLATNLASEA 660

QY      661  SDKGNLPKYSHNSLMVQAIKTNLITDPDIHVLFPDVEALIQLLITEASRPNTALISPENL 720
DB      661  SDKGNLPKYSHNSLMVQAIKTNLITDPDIHVLFPDVEALIQLLITEASRPNTALISPENL 720

QY      721  QKASGSSDKGGSFLTQKRAAVLFQVKETIKENIKHLLDDEDEEIEIMRQREESDPEY 780
DB      721  QKASGSSDKGGSFLTQKRAAVLFQVKETIKENIKHLLDDEDEEIEIMRQREESDPEY 780

QY      781  RRSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSPGLLSR 840
DB      781  RRSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSPGLLSR 840

QY      841  GGHMSLMLPGYNOPACKLSHGKTEVGKRLPASEGVCKGTGVSVRAVTTQHLLSIISLANT 900
DB      841  GGHMSLMLPGYNOPACKLSHGKTEVGKRLPASEGVCKGTGVSVRAVTTQHLLSIISLANT 900

QY      901  LMSMTNATFIGDHMKKGTPRPPRPSTPDLKARGSPPTSSNIVQGOIKQVAAPVVSARS 960
DB      901  LMSMTNATFIGDHMKKGTPRPPRPSTPDLKARGSPPTSSNIVQGOIKQVAAPVVSARS 960

QY      961  ADHSGSDPPSAPALHTCFLVNEGWSQLAAHVCMLPDLGLDKFRPILMLARRQDRC 1020
DB      961  ADHSGSDPPSAPALHTCFLVNEGWSQLAAHVCMLPDLGLDKFRPILMLARRQDRC 1020

QY      1021  LEVREAAQALLAELRLIEQAGRKRAIDAWAPVLPQYIDHVISPGVTSAAQTITAPDA 1080
DB      1021  LEVREAAQALLAELRLIEQAGRKRAIDAWAPVLPQYIDHVISPGVTSAAQTITAPDA 1080

QY      1081  SGPEAKVQSEEHDLVDDDDITTCGLSSVPQMKKISTSYEERRKQATAVLLGVIGAEFGAE 1140
DB      1081  SGPEAKVQSEEHDLVDDDDITTCGLSSVPQMKKISTSYEERRKQATAVLLGVIGAEFGAE 1140

QY      1141  IBPPKLLTPRSSSQIPEGFGLTSGGSNYSLARHTCKALTFLLQLPPSPKLPHPSTIRRT 1200
DB      1141  IBPPKLLTPRSSSQIPEGFGLTSGGSNYSLARHTCKALTFLLQLPPSPKLPHPSTIRRT 1200

QY      1201  AIDLIGRGFTVPEYMDVSAVLMLLELCADEKQALANTMGLPLSPAADSARSARHALS 1260
DB      1201  AIDLIGRGFTVPEYMDVSAVLMLLELCADEKQALANTMGLPLSPAADSARSARHALS 1260

QY      1261  LIATARPPAFITTIKAEVHRHTALAAANTOSQOQNMHTTTLARAKEILRVIELIEKMPD 1320
DB      1261  LIATARPPAFITTIKAEVHRHTALAAANTOSQOQNMHTTTLARAKEILRVIELIEKMPD 1320

QY      1321  VVDLLVEVMDIIMYCLEGLSVKKKGLQECFPAICRPFYMYVYERNHRIAVGARHGSVALY 1380
DB      1321  VVDLLVEVMDIIMYCLEGLSVKKKGLQECFPAICRPFYMYVYERNHRIAVGARHGSVALY 1380

QY      1381  DIRTGKQPIHGHKGPIITAVAPDGRYLATYNTDISHISFWQMTSLGSGTGLNSAPQ 1440
DB      1381  DIRTGKQPIHGHKGPIITAVAPDGRYLATYNTDISHISFWQMTSLGSGTGLNSAPQ 1440

QY      1441  LRCKITQYVPPVQAPSPGSHNALKLARLWTSNRNVLMAHDKGKEHFRFW 1490
DB      1441  LRCKITQYVPPVQAPSPGSHNALKLARLWTSNRNVLMAHDKGKEHFRFW 1490
```

RESULT 3

```
US-10-094-749-2956
; Sequence 2956, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
```

APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOKIHO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2956
LENGTH: 559
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2956

Query Match 36.1%; Score 2805.5; DB 15; Length 559;
Best Local Similarity 94.3%; Pred. No. 1.1e-229;
Matches 549; Conservative 0; Mismatches 0; Indels 33; Gaps 1;
QY 526 SARVQHICVADSHVGLLSREKKCIMLASRHLPIQVWKRPDDVYVGCSDGSVY 585
DB 11 SARVQHICVADSHVGLLSREKKCIMLASRHLPIQVWKRPDDVYVGCSDGSVY 70
QY 586 VQWMDTGALDRVCWVGITAVEILNACDEAVPAADVSLSHPAVNLQAMTRRSALAKNMAH 645
DB 71 VQWMDTGALDRVCWVGITAVEILNACDEAVPAADVSLSHPAVNLQAMTRRSALAKNMAH 130
QY 646 HKLQTLATNLLASEADSGNLPKYSNLSMVQAIKTNLTDPDTHVLFFDVEALIIQLLTE 705
DB 131 HKLQTLATNLLASEADSGNLPKYSNLSMVQAIKTNLTDPDTHVLFFDVEALIIQLLTE 190
QY 706 EASRPNTALISPENLQKAGSSDKGGSFLTGGKAAVLFQVQKTIKENIKHLLDDEED 765
DB 191 EASRPNTALISPENLQKAGSSDKGGSFLTGGKAAVLFQVQKTIKENIKHLLDDEED 250
QY 766 EETMRORRESDPEYRSSKPLTLLLEYNLTMDTAKLFMSCLHAWGLNEVLDVCLDRLG 825
DB 251 EETMRORRESDPEYRSSKPLTLLLEYNLTMDTAKLFMSCLHAWGLNEVLDVCLDRLG 310
QY 826 MLKPHCTVSGLLSRGHHMSLMPGVNQPACKLSHGKTEVGRKLPASEGVGKGTGVGSRA 885
DB 311 MLKPHCTVSGLLSRGHHMSLMPGVNQPACKLSHGKTEVGRKLPASEGVGKGTGVGSRA 370
QY 886 VTTHQLLSIIISLANTLMSMTNATFIDHMKKGTPRPRSTPDLSTARGSPPTSSNIVQ 945
DB 371 VTTHQLLSIIISLANTLMSMTNATFIDHMKKGTPRPRSTPDLSTARGSPPTSSNIVQ 430
QY 946 QIKQVAPVVSARSADSHSGSDPPSPALHTCFLVNEGWSQLAAMHCWMLPDLGLDKFR 1005
DB 431 QIKQ-----GWSQLAAMHCWMLPDLGLDKFR 457
QY 1006 PPLEMLARWQRCLEVREAAQALLAEARRIEQAGRKEAIDAWAPYLPQYIDHVISPG 1065
DB 458 PPLEMLARWQRCLEVREAAQALLAEARRIEQAGRKEAIDAWAPYLPQYIDHVISPG 517
QY 1066 VTSEAAQTITAPDASGPEAKVQEEHDLVDDITTCGLSSV 1107
DB 518 VTSEAAQTITAPDASGPEAKVQEEHDLVDDITTCGLSSV 559

US-10-108-260A-3494
Sequence 3494, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: No. US20040005560A1el full length cDNA
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3494
LENGTH: 775
TYPE: PRT
ORGANISM: Homo sapiens
US-10-108-260A-3494

Query Match 18.6%; Score 1442; DB 15; Length 775;
Best Local Similarity 38.6%; Pred. No. 4.5e-113;
Matches 303; Conservative 173; Mismatches 268; Indels 40; Gaps 11;
QY 10 IVLWGRKAPTHCISAVLLTDDGATVTCGCHDQICLWDLSELQINPRALLFGHTASITC 69
DB 8 VALWGQKAPPHSITAITMITDDQRTTVTGSQEQQLCWNLSHELKISAKELLFGHSASVTC 67
QY 70 LSKACASSDKQYIVSASESGEMCLWDVSDGRICEFTKLTACTHTGTQFYQFSVGNQREGRL 129
DB 68 LARARDFSKQYIVSARENGEMCVNVTNGQCVKATLPYRHTAICYHCSFRMTGEGWL 127
QY 130 LCHGHPYELVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQBDTVVALSVTGILKVM 189
DB 128 LCGEYQDVLIIDAKTLAVVHFRSSQPPEDWNCNCIVHSMRQEDSLVWVAVAGELKVM 187
QY 190 IVTSEISMDQTEPIFEESKPIYCONCOSISFCAPTQBSLLVVCCKWYRVDAGDYSLL 249
DB 188 DLSSINSIQERQDQVYKESKLESLNCQTIREFCTYTERLLLVVFSKCKWVDYCDLSLL 247
QY 250 CSGPENGQWVGDPVSSDKVLIWTEGOSYIYKLPASCLPASDSFRSDVCKAVENLTP 309
DB 248 LTVERNQOFFAGGEVIAHRLIWTEDGHSYIYQLLSGL--SKSIYPADGRVLKETIY 305
QY 310 PVQHIILD---RKDKELLICPPVTRFPYGCREFYFKLLIQGSDSGRLNIWNISD--TADK 364
DB 306 P--HLLCSTSVQENKE-QSRFPVMGMNERKEPFYKVLFSGEVSGKITLWHIPDPVPSKF 362
QY 365 QGSEGLAMTTSISIQEAPDKNCPAGIIDLQSVIPNSNEPLKTASVYIIPAHGELVCG 424
DB 363 DGSPREIPVTATWTLQDNFQKHTMSQSIIDYFSGLKDGAGTAVVTSSEYIIPSLDKLIG 422
QY 425 REDGSIVIVPATQTAIVQLQGEHMLRGMPPHRTLGRHNRKVTCLLYPHOVSARYDQRY 484
DB 423 CEDGTIIITQALNAKARLEGGSLVKDS-PHKVLKHHQSVTSLLYPHGLSSKLDQSW 481
QY 485 LISGVDPSVIIWIDIFSGEMKHI FCVHGGEITQLLVPPENCARSARVQHCTCSVASHSVGL 544
DB 482 MLSGDLSDCVILWDIFTEILHKFLEAGPVTSLMSPEKFKLRGEQIICCVCGDHSVAL 541
QY 545 LSIREKKCIMLASRHLFPQVTKWRPSDDYLVGSDGSVYVQWMDTGALDRVCWGITAV 604
DB 542 LHLEKSCCLLHARKHLFPVRMIKHPVENFLVGCADDSVYIWEIETGTLEHETGERAR 601
QY 605 ETILNACDEA-----VPAADVSLSHPAVNLQAMTRRSALAKNMAHKLQTLATNLLAS 658
DB 602 IILNCCDDQLVKSVLPITASETLKHKSIQRSS-----SPYQGLPCLPCLQV 649
QY 659 EASDKGNLPKYSNLSMVQAIKTNLTDPDTHVLFFDVEALIIQLLTEAS--RPNTALIS 716
DB 650 ESSCKVTDKFCPRPENVLPVTKWNSVGHILLFDLENLVELLPTPLSDVDSSSSFFYG 709
QY 717 PENLQKAGSSDKGGSFLTGGKAA-----VLFOQVQKTIKENIKHLLDDEEDDEEIM 769
DB 710 GEVLRRAKSTVBEKKTILTLFKSKTACGPLSAEALAKPITESLAQG--DNTIKFSEENDGIX 767

QY 770 RQR 773
Db 768 RQK 771

RESULT 5
US-10-437-963-162857
; Sequence 162857, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162857
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61908C.1.pap
US-10-437-963-162857

Query Match 5.3%; Score 412.5; DB 16; Length 1391;
Best Local Similarity 19.5%; Pred. No. 8.9e-25;
Matches 340; Conservative 210; Mismatches 560; Indels 631; Gaps 81;

QY 12 LWRKAPTHCISAVLLTDDGATVTCGCHQICLWLSVBLQINPR--ALLFGHTASIT- 68
Db 9 LWSPPSPSHITAAATP--AALFTGAADGTILHWPLPPSPRSPSSILLCAHAAITS 66
QY 69 -----CLSKACASSDKQYIVSASEGEMCLWVSDGRCIEFTKL----- 107
Db 67 LCPLSPSPCLLASCAGVLS-LFSSSASASL-----RCLRRSLPWAQSPSLVAP 119
QY 108 ---ACTHTGTFQYQFVSGNOREGLCHG-----HYPEILVVDATSLEVL-----YSLVS 154
Db 120 LPSSSSAG-----SSASVAILCHAPDDGGRHVSAAVVVDARTLVLRATFGALS 171
QY 155 KISPDWI-----SSMSIIRSHRTQEDTVVAL-----SVTGILKVVITVSEISDMQD 200
Db 172 VAPPRAIAVADAGVEDASVSVLADAGQRAQVVPVVAEGAAGVSDGSPRRLSASSASVTS 231
QY 201 TEPI-----PEESKPIYQNCQSI SFCATQBSLAVVCSKYRVPFDAGDYS-----L 248
Db 232 AEAVDGRVAVSLSDGKV-----ALVMKNSCLLKISEGVWL--GSEVTLPSDL 279
QY 249 LCSPSPSNGQW-TGGDFVSSDK-----VIWTENGOSIYKLPASCLPAS 293
Db 280 LCKEENGKMGVLGVGFFLRGGENGAGHSENGNVRSVLVMSINGAIVY----- 329
QY 294 DSRSDV-----KAVENLIPVQHILLDRDKELL-ICPPVTRPFYGCYFHKLLIQ 346
Db 330 ---RVEVGTSGFGKAV---CBIPDIVSBERGDSLVQFCQGNQ-----LIR 370
QY 347 GDS-----SRL-----NTW-----NISDTADK-----QSEBGLAMTTSISLO 380
Db 371 VESRPYKIASLLWKPFVSIWSDHLELNTANNIERPPLSKILGEGLOGEERSDHS-- 428
QY 381 EAFDKLNPCCAGI-IDOLSVIPNSN-----EPLKVTSVYTPAHGRVCGREDGS 429
Db 429 HSFQSN---NGVDINSLICSSNSGLGRGGTVSSMWLSSESYTPY--AVVYGFHGD 483
QY 430 IVIVPATQTAIVQLQOEHMLRRGWPPH---RTLGRHNRKVTCLLYPH---QVSARYDOR 483

Db 484 IEVI-----RFLNLLPAKFGSGGIYPHISERFFLGTGAILCLAAHMHMAOPDSRTFNR 538
QY 484 YLISGGVDFSVIWDIFSGEMKHIFCVHGGEITQLQLPPENCARSARVQHICISVASDHSV 543
Db 539 VLISGSFDSITRVWDLDACTILSVMHHPVAPVKQIMLPANWTHQPDCCFLSVGEIGIVA 598
QY 544 LLSREKKCICIMLASRHLFPQVTKWRPSDDYLV-----GCSG-SSV-YVWQMDTGALD 595
Db 599 LNSLQTRVERMPFGHPSPYPMVMDGVKGYIACLRNLHSCNDSGLVYIMDLTKTGARE 658
QY 596 RCVMGITAIVELL-----NACDEAVPAADVLSHPAVNL-KQATRRSLAALXOMA- 644
Db 659 RIITGTSOSSTFEHFCRGISKNAVTGSIIGTTSASSLLVPIPKOTSLLOSHANKGLSI 718
QY 645 -----HHKLTQTLATNLASEASD--KGNLPKYSH-----NSLMVQAI 679
Db 719 SSVSTVNHANVTNSVTPASADVGMKMSATDEAHLHGNSGKVASGQCINNRRKPI 778
QY 680 KTNLTDPDIHVLFFDVEALIQLLBEASRPNTALISPENLQKASGSDKGGSLTQKRA 739
Db 779 KCSCPYPGIASLRFDL-----TAIMSTQGM--ANNNSDR----- 810
QY 740 AVLFOQVKETIKENIKHEHLLDDEEIEIMRQREES-----DREYSSKSKPILTLE 792
Db 811 -----QLRDHFYRDVND-SIQAETCDNTSGMHVIDSPRESLEGR---LLR 853
QY 793 YNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPH-CTVSFGLSLRGHSLMLPGY 851
Db 854 FSLCP-----LHLGVDELKLDVDEMQVCEKCHTATGVVGRGSLTFMPPG- 903
QY 852 NOPACKLSHGKTEVGRKLPASEGVGKTYGVSAVTTQHLLSIIS-LANTLM-SMTNATP 909
Db 904 -----KEATLEIWSSSAKCLMDNQVNAIL 927
QY 910 IGDHMKKGTRPRRSTPD-----LSKARGSPPTSSNIVOGQIKQVAPVVSARSADHSG 965
Db 928 MDLQLLKNKVPDSQLFTSQDMNIIITAIQASVSS---YQOLKADNEDV--GREDCDTSE 982
QY 966 SPPSPAPALHTCFVLNVEGWSQAAHMCVMLPDLGLDKF-RPPLLEMLARRWDRCLVR 1024
Db 983 I-----SSW-----LESFENQEWLSWIGTSQD----- 1005
QY 1025 EAAQALLAELRRIEQAQKEAIDAWAP-----YLPQYIDHVISPG--VTSEAAQ 1072
Db 1006 AVASNIIVA-----AALVWVPSIVKPKLAHLVNVQLIKLWMSNDRYSSTAEE 1054
QY 1073 TTTAPPDASGPAKQVEEHLVDDDDITTCGLSSVPQMKKISTSYERRKQATAIVLLGV 1132
Db 1055 LLAEGMESTWKVCLGDMTHFLSDVLFOIECLSSAPSNNAV-----YKTAVALTMR-- 1105
QY 1133 IGAFFGAETEPKLLTRPRSSQIPGFGLTSGGSNYSLARHTCKALTFLLLQPPSPKLP 1192
Db 1106 -----EALVGTLL-----PSL- 1116
QY 1193 PSHITRTAIDLIGRGF-----TWPEYMDVSAVLMGLLELCADAELKOLANITMGLP-- 1244
Db 1117 -----AMADIV--GFFGVIOQIATSSDSVPVHVISL-----KTLIRVVRSPKA 1159
QY 1245 LSPADSARSARHALSLIATARPAPITTIKAEVHRHTALAAANTOSQONHHTTLARAKG 1304
Db 1160 LAPYLDK-----AISV-----LHT----- 1174
QY 1305 EILRVIEILLERKPTDVVLLVEMDITM--YCLBGLSVKKKGLQECFPAICRFYMVSY 1362
Db 1175 -----MDPSNLMRKACIINSMALREIARVFP-----MVALN 1207
QY 1363 ERNHRIAVG-----ARHGSVALYDITGKQTIHGKGP-----ITA 1399
Db 1208 ESMTRLAVGDAIGEHNATIRVYDIESVTKIRILDASGPPGLPSLLDGSNTTATILITA 1267
QY 1400 VAFAPDGRYLATYSNTDISHISFWQNTSLLSGISGMLNSAPQRCIKTYQVPPVQSPASPS 1459
Db 1400 VAFAPDGRYLATYSNTDISHISFWQNTSLLSGISGMLNSAPQRCIKTYQVPPVQSPASPS 1459

Db 1268 LSFLEGEGLVAFSENGLMIRWNSLGNWWRLS--RSLTPIQCTKLIYVPPWEGFSPNS 1325
Qy 1460 -----HNAKLARLIWTSNRNVILMAHDKG 1484
Db 1326 ARLSIISILGHDKHONSETKTRELDEADNLKLLAHNLDSLRYLQWVGKTIKLTTRH-GQ 1384
Qy 1485 E 1485
Db 1385 E 1385

RESULT 6
US-10-369-493-18893
; Sequence 18893, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18893
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(579)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-18893

Query Match 2.9%; Score 228.5; DB 15; Length 579;
Best Local Similarity 20.6%; Pred. No. 9.9e-10;
Matches 120; Conservative 80; Mismatches 203; Indels 179; Gaps 24;

Qy 19 TCISAVLLTDDGATIVTGCHDQICLWDLV-ELQINPRALLFGHTASITCLSKACASSD 78
Db 109 TSCVRSVVFSDGAMLASGSDDTVRLWDIS---SGNCLYTLQHT---SCVRSVVFSPD 162
Qy 79 KQIVSASEGEMCLWDVSDGRCIEFTKLACTHTGTFQYQFVSGNOREGRLCHGYPEI 138
Db 163 GAMLASGDDQIVRLWDISSGNCL-YTLOGYT-SWVRFLVFS---PNGVTLANGSSDQI 216
Qy 139 L-VVDATSLVLSVLSKISPDWISSMSIIRSHRTQEDTVVALSVTG--ILKVIWTVTSEI 195
Db 217 VRLWDISSKKCLYTLQ--HTNWNNAVAF-----SPDGATLASGSDQTVRLWDISS-- 266
Qy 196 SDMDQTEPIEESKPIYQNCOSISFCATQRLSLLVCSKYW---RVPDAGYSLLCG 252
Db 267 -----SKLY-----ILQHTSWNSVVFNP-DGSTLASG 295
Qy 253 PSENGQTWGTGDFVSSDKVI-IWTENGQSYIYKLPASCLPASDSFRSDVKAVENLIPPV 311
Db 296 -----SSDQTVRLWEINS-----SKCLCTFQHTSWNSVVFN----- 328
Qy 312 QHILLDRKOKELLICPPVTRFFYGCREFYHKLLIQDSSGRLLNIWISDTADKQSGEGL 371
Db 329 -----PD-----GSMLASGSSDKTVRLWDISSSKCLHTTQGH 361
Qy 372 AMTTSISLQEAFLKLNPCPAGIIDQLSVIPNSNEPLKVTASVIPAHLGLVCGREGSIV 431
Db 362 NWVNSVA-----FNP-----DQSM 376
Qy 432 IV-PATQTAIVQLLQGEHMLRRGWPPHRTLRGHRNKVTCLLYPHQVSARYDORYLISGGV 490

Db 377 ASGSGDQTVRLWEISSKCL-----HTFQHTSWVS-----SVTFSPDGTMLASGSD 423
Qy 491 DFSVLIWDIFSGEMKHIFCVHGGEITQLLAVPPENCARSVQHCHICSVASDHSVGLLSLREK 550
Db 424 DQTVRLWISSGECLYTFGLHTNWGVSIVFSPDGA-----ILASGSDQTVRLWISSG 477
Qy 551 KCIMLASRHLFPPIQVIKWRPSDDYLWVGSDGVSVVWQMDTG 592
Db 478 KCLYTLQGHNNWVGSI VFPSDGTLLASGSDDTVRLWNISSG 519

RESULT 7
US-10-369-493-20224
; Sequence 20224, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20224
; LENGTH: 608
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme
US-10-369-493-20224

Query Match 2.9%; Score 228.5; DB 15; Length 608;
Best Local Similarity 18.4%; Pred. No. 1.1e-09;
Matches 112; Conservative 89; Mismatches 176; Indels 231; Gaps 25;

Qy 23 SAVLLTDDGATIVTGCHDQICLWDLV-ELQINPRALLFGHTASITCLSKACASSDKQY 81
Db 27 AGIAPSPDFTLLATDAGELRLEWATGKLVN-----FAGHLGWVWSLA---FSPDQGL 79
Qy 82 IVSASBSGEMCLWDVSDGRCIEFTKLACTHTGTFQYQFVSGNOREGRLI----- 130
Db 80 LASCSSDKTIRLWDVNTGKCLR-----TLSGHTSIWSVAFSADQOMLASGDEPTIRL 133
Qy 131 -----CH-----GHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVA 179
Db 134 MNVNTGDCHKIFSGHTDRILSL-----SFGSDGQTLAS 166
Qy 180 LSVTGILKVIWTVTSEISDMQDTEPIEESKPIYQNCOSISFCATQRLSLLVVCYSKYR 239
Db 167 GSADFTIRLWKISGE-----CDRILEGHSRDIW-----SISF----- 198
Qy 240 VPDAGDYSLLCSPSNGQWTGDFVSSDKVI-IWTENGQSYIYKLPASCLPASDSFRS 298
Db 199 -----SPDQGLVSG---SADFTIRLWEVS-----TGNCF----- 225
Qy 299 DVGKAVENLIPVQHILLDRKD--KELLI CPPVTRFFYGCREFYHKLLIQDSSGRLLNIW 356
Db 226 -----NILQHSDRVRSALAFSPA-----QMLVSASDDKTIRIW 259
Qy 357 NISDTRADKQSGSEGLAMTTSISLQEAFLKLNPCPAGIIDQLSVIP-NSNEPLKVTASVVI 415
Db 260 EAS-----TGEC-----LNLPGHTSIWSVAFVFN-- 284
Qy 416 PAHGLRVCGREGDSIVIPATQTAIVQLLQGEHMLRRGWPPH-----RTLGRHNRKVTCL 470
Db 285 -----DQRTTASGSDQTV-----KLDVNTGRCFKTLKGYNSV--- 319
Qy 471 LYPHQVSARYDORYLISGGVDFSVIWDIFSGEMKHI FCVHGGEITQLLVPENCARSVQ 530

Db 320 ---FSVAFNLDGQTILASGSDTQTVRLWDVNTGTCCKFAGHSGWTVSVAFHPDG----- 370

Qy 531 HCICSVASDSVGLSLREKCKIMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSGVYVQMD 590

Db 371 DLLASSADRTIRLWSYSTGCLQILKDHVNWQSVAFSPDRQILASGSDDTIRLWSVS 430

Qy 591 TGAIDRCV 598

Db 431 TG---KCL 435

RESULT 8

US-10-425-114-58817

; Sequence 58817, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 58817

; LENGTH: 703

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: 700086365_FLI.pep

US-10-425-114-58817

Query Match 2.9%; Score 228; DB 15; Length 703;

Best Local Similarity 17.9%; Pred. No. 1.5e-09;

Matches 166; Conservative 126; Mismatches 273; Indels 360; Gaps 33;

Qy 671 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 730

Db 16 HNNKIKCPKICSPYGIASLRFDL-----TAIMSTQGM--TNSNDR- 56

Qy 731 GSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEEIMR-----QRRESDPEYRSS 783

Db 57 -----HLKDHLL--HSENPKEMLQPGTLDSPSRVHENDSPSR 92

Qy 784 KSKPLTLLRYNTMTDTAKLFMSCLHAWGLNEVLDEVCLEDRGLMKLPH-CTVSPGLSRGG 842

Db 93 LEG--CLLRFSLCF-----LHLWDVDCLEKLLVDEMVCCKPEGCHIATGVVDKG 141

Qy 843 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 902

Db 142 SFTLMPFG-----KEATLELWKS-----SABFCAMRSLSIISLAQLMI 179

Qy 903 SMTNATPIGDHMKKGTPRPRSPDLSKARGSPPTSSNIVQGIQVAAPVVSARSAD 962

Db 180 KLSRSC-----TNSN----- 189

Qy 963 HSGSDPPSAPALHTCFLVNEGWSQLAMH---CVMLPDLLGLDKFRPPLLEMLARWQD 1018

Db 190 -----SALAAYTRHFAEKVPDI-----KPPSLQLLVSWFQW 221

Qy 1019 RCLVREAAQALI-LAELRBIQAGRKEADAWAPYLPOVIDHVPSTSE----- 1069

Db 222 PSEHVRMARSLFHCAAPRSIPQPLRVNK-----NKVNPAPLSSDDMMNVLIC 269

Qy 1070 AATITITPADSGPEAKVQEEHDLVDDITTCGLSSVPQMKKISTSYEERRQATAVL 1129

Db 270 AVGTASVS-----NHGELKEDDENVDKDDIDT-----ANWILM 302

Qy 1130 LGVIGAEFGAIEBPPKLLTRPRSSQITPEGFLTSGGSNYSLARHTCKALTFLLLQPPSP 1189

Db 303 L-----ESFENQEWLSWIGTSDAVASNIIVAAALVVMYPSV 341

Qy 1190 KLPPHTIRRTAIDLIGRGFTVWEPYMDVSAVLMGLLELCADAEKOLANTMG----- 1242

Db 342 KAKLSLVVSQILKV---MSMDRYSSTAA-----ELLAEGMENTWAKACAEITHFM 392

Qy 1243 -----LPLSPAADSARSARHALSL-----IATARPPAFITTIKAEVHRHT 1282

Db 393 SDILFOIECLSTAPSSNAINKTAIVATMREALVGTLLPSLAMADVGTGFGVIESQIWA-- 450

Qy 1283 ALAANTOSQONMHTTTLARAKGEILRVIELIEKEMPTDVVDLLVVEWMDIMY----- 1334

Db 451 -----TSSDSPVHVASIK---TIIRV---VRGAPKSLVPYLDKAINYTLHTWDPNSLI 497

Qy 1335 -----CLEGSLVKKKQEGCFPAICRFVMSVSYERNHRIAVG-----ARHGSVALYDRTG 1385

Db 498 MRKACIISSMMALREMARVFP-----MWALNESMTKLAVGDAIGEYDTIRVYDIESV 551

Qy 1386 KCOITHGKGP-----ITAVAPADGRYLATYNTSDSHISFWMQNTS 1427

Db 552 TKIRILDASGPPGLPSLLAGSSNRTTILISVLSFSPDGBGLVAFSENGLMIRWWSLGS 611

Qy 1428 LLGSIGMNSAPQLRCIKTYQVPPQASPGS----- 1459

Db 612 WWERLS--RSLTPIQCTKLIYVPPWEGFSPNSARLSIISILDKPKRSEKELGEADNLK 669

Qy 1460 ---HNALKLARLIWTSNRNVILMAH 1481

Db 670 LLLHNDLSYRLHWIGGKTIKLRH 694

RESULT 9

US-10-425-115-360670

; Sequence 360670, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 360670

; LENGTH: 710

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577_92112C.1.pep

US-10-425-115-360670

Query Match 2.9%; Score 228; DB 16; Length 710;

Best Local Similarity 17.9%; Pred. No. 1.5e-09;

Matches 166; Conservative 126; Mismatches 273; Indels 360; Gaps 33;

Qy 671 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 730

Db 23 HNNKIKCPKICSPYGIASLRFDL-----TAIMSTQGM--TNSNDR- 63

Qy 731 GSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEEIMR-----QRRESDPEYRSS 783

Db 64 -----HLKDHLL--HSENPKEMLQPGTLDSPSRVHENDSPSR 99

Qy 784 KSKPLTLLRYNTMTDTAKLFMSCLHAWGLNEVLDEVCLEDRGLMKLPH-CTVSPGLSRGG 842

Db 100 LEG--CLLRFSLCF-----LHLWDVDCLEKLLVDEMVCCKPEGCHIATGVVDKG 148

Qy 843 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 902

Db 149 SFTLMFPG-----KEATLELWKS-----SABFCAMRSLSIISLAQLMI 186
Qy 903 SMTNATFIGDHMKGPTRPRPTDLSKARGSPPTSSNIVQOIKVAAPVVSARSAD 962
Db 187 KLSRSC-----TNS-----196
Qy 963 HSGSDPPSAPALHTCFLVNEGWSQLAAMH-----CVMLPDLGLDKFRPPPLLEMLARRWQD 1018
Db 197 -----SALAIFYTRHFAEKVPDI-----KPSLQLLVFWQH 228
Qy 1019 RCLVETRAQAQLL-LAELRRIEQAGRKEADAWAPYLPQVIDHVISPGVTSE-----1069
Db 229 PSEHVRMAARSLFHCAPRSIPQPLRVNK-----NKVVNAPLSSSDNMNVLIC 276
Qy 1070 AAOITITAPDASPEAKVQEEHDLVDDDITTCCLSSVPMKKISISYEBERRKQATAVL 1129
Db 277 AVQTASVS-----NHGELKEDDENVDKDDIT-----ANMILW 309
Qy 1130 LGVIGAEFGAIEBPPLKLLTRPRSSQIPBEGFLTSGGSNYSLARHTCKALTFFLLQPPSP 1189
Db 310 L-----ESFENQEWLSWIGTSQDAVASNIIVAAALVWVYPSV 348
Qy 1190 KLPHPSTIRTAIDLGRFTWEPYMDVSAVLMGLLELCADAEKQOLANITWG-----1242
Db 349 KAKLSLIVWSQLIKLV---MSMNDRYSSTAA-----ELLAEGMENTWKACLGAEITHEW 399
Qy 1243 -----LPLSPAADSARSARHALSL-----IATAPPAFITTIAKEVHRHT 1282
Db 400 SDILFQTECLSTAPSSNAINKTAVAITMRBALVGTLLPSLAMADVTFGFGVIESQIWA- 457
Qy 1283 ALAANTOSQONMHTTTLARAKGEILRAKIEILBKMPDVVDLIVEVMDIIMY-----1334
Db 458 -----TSSDSPVVASIK-----TIIHV-----VRGAPKSLVPLDKAINILHTMDPSNLI 504
Qy 1335 -----CLEGSUVKKKGLOECPAICRFPMVSYVYERNHRIAVG-----ARHGSVALYDIRTG 1385
Db 505 MRKACIISMMALREMARVFP-----MVALNESMTKLAVGDAIGEYDITRVYDIESV 558
Qy 1386 KCOTIHCHKGP-----ITAVAPADGRLYATYNTDISHISFQMNTS 1427
Db 559 TKRIILDAGPGLPGLSLLAGSSNRTTILISVLVSFSPDGEGLVAFSENGMLMRWWSLGS 618
Qy 1428 LLGSIGMNSAPQLRCIKTYQVPPVOPASPGS-----1459
Db 619 WHERLS--RSLTIQCTKLIYVFPWEGFSPNSARLSIISILLDDKPKRSEKELGEADNLK 676
Qy 1460 ---HNALKLARLIWTSNRNVILMAH 1481
Db 677 LLLHNLDSYRLHWIGGKTIKLTTRH 701

RESULT 10
US-10-369-493-20242
; Sequence 20242, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20242
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme

US-10-369-493-20242
Query Match 2.8%; Score 220.5; DB 15; Length 1005;
Best Local Similarity 20.3%; Pred. No. 1.2e-08;
Matches 124; Conservative 93; Mismatches 251; Indels 143; Gaps 24;
Qy 22 ISAVLLTDDGATIVTGCHDGOICLDLSVELQINPRALLFHTASITCLSKACASSDKQY 81
Db 465 IRSIAYSPNGQQLVSASADTKIKIWDVSSGKLLK---TLTGHTSAV---SSVAYNPNGQ 518
Qy 82 IVSASESGEMCLWDVSDGRCIEFTKLACTHTGQIQYQVSGVGNOREGRLLCHGHPY-ILV 140
Db 519 LASASDNTIKIWDISSG-----KLLKTLPLGHSSVSVSNVAYNPNGQQLASANDKTIKI 572
Qy 141 VDATSLEVLYSVKISPDWISSMSIIRSHRTQEDTVVALSVTGILKWIVTSEISDMQD 200
Db 573 WDINSGKLLKSLTGHS-----EVNSVAYSPNGQ---LASASFDNTIKIWDISS--GKLJK 624
Qy 201 TEPIFEESKPIYCONCOSISFCAPTQSRLLLVVCSKYMRVFD-----AGDYSLLC 250
Db 625 T---LTGHSNVVP-----SVAYSPNGQHLASASADTKIKIWDVSSGKPLKSLAGHSNVVP 676
Qy 251 S-GPSENGGTWGTGDFVSSDKVI-IW-TENGQSYTYKLPASCLPASDSFRSDVGKAVEM 307
Db 677 SVAYSPNGQQLAS---ASDDKTIKIVMDISNG---KPLESMTDHSRNVSV-----Y 722
Qy 308 IPPVOHILLDRDKELLICPPVTRFYGCYREYFHKLLIQGSSGRININISDTADKQGS 367
Db 723 SPNGQHLASPSYDK-----TIKIWNVS-----744
Qy 368 BEGLAMTTSISLQEAFDKLNPCAGIIDLQSVIP-----NSNEPLK-----V 409
Db 745 -SGKLLKTLTGHSSESVNSVAYSPNG--QQLASASWDKTIKWWDVNSGKPLKTLIGHSSV 801
Qy 410 TASVYIPAHGRVLCGREGDSIVIVPATQTAIVOLLOGEHMLRRGPHPTLGHNRKVT 469
Db 802 NSVAYSPNGQQLASASFDNTIKIWDVSSGKLL-----KTLTGHSNAVS 845
Qy 470 LLYPHQVSARYDORYLISGGVDFSVIWDIFSGEMKHIFPCVHGGEITOLLVPPENC 529
Db 846 VAYSP-----NGQQLASASLONTIKIWDVSSAKLLKTLTGHSDAVSSVAYSPNG-----894
Qy 530 QHCISVASDHSVGLLSREKKCIIMLASRHLFPFIQVIKWRPDDYLVVGCSDGSVYVWQM 589
Db 895 -QQLASASDNTIKIWDVSSGKLLKSLGHSNAVSIAVSPNGQQLASASADNTIKIWDV 953
Qy 590 DTGALDRCVMG 600
Db 954 SSGKLLKSLG 964

RESULT 11
US-10-369-493-20247
; Sequence 20247, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20247
; LENGTH: 626
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme

US-10-369-493-20247

Query Match 2.6%; Score 205.5; DB 15; Length 626;
Best Local Similarity 17.4%; Pred. No. 1e-07;
Matches 124; Conservative 116; Mismatches 244; Indels 229; Gaps 31;

QY 4 NSLVLPV-----LWGRKATHCISAVLLTDDGATVTCGHCQICLWDL 48
DB 15 NQLAPVIGSLKAMNTPTRANSURGHEQVNCVA---FSPDGKFTASGSDSTLCLWNI 71
QY 49 SVELQINPRA-LLFGHTASITCLSKACASSDKQIVTSASBSGEMCLWDVSDGRCIBFTKL 107
DB 72 IG---NPTAQFLCGHEQVNCIA---FSPDGKFIASGSDIGLCLWDL-----113
QY 108 ACTHTGQVQFVSGVGNRGRLLC---HGYPEILVV-----DATSLEVLYSLVS-----154
DB 114 -----QGNLTQPMQGHGEEGVISVAFSPNSDGCANPGSVISVGF 155
QY 155 -----KISPDWISSMSIIRSHR-----TQEDTVVALSVTGLKVIWITSE 194
DB 156 GTVCLMDLQNALITQPW-----RGHKEGVISVAFSPNGDCIISVGFDTCLW---D 204
QY 195 ISDMQDTEPIFEESKPIYQNCQSIISFCAPTQSRLLVVCYSKYRVPDAGDYSLLCSGGS 254
DB 205 LEGNTITQPMHKEAK-----IICATF-----SPDRKFIYVSGS 238
QY 255 EN-----GQWTG-----GDFV---SSDKVI-IWTEGQSYIYKL 285
DB 239 DSTVRLWDIQGNPIGQPMHGHGVNSVAFSPDGKFTISGCDTIRLWNINGS-----293
QY 286 PASCLPASDFRDVGKAVENLPPVQHILLDRKXELLICPPVTRFFYGCRRYFKLL- 344
DB 294 -----ITQWRGHEGVNSLAFSPDGKLIISGDRTVRLWE-----LHQILQ 335
QY 345 --IQDSSGRINLWNTSDTADKQSBEGLAMTTSISLQEAFLDKLPCPAGILQLSVIPN 402
DB 336 DRVIGRSQRYENWNSVAFSPDG---QWIVSASNDSTIRLWDS-NGNPTG-----QW 395
QY 403 SNEPLKVTASVYIPAHGRVLVCGREDGSIIVPATQTAIVOLLOGEHMLRRGPPHRTLRG 462
DB 386 QGHEKEVNSVAFSP-----DQO-QWIVSASNDSTIRLWDS-----NGNPIGQPMQ 429
QY 463 HRNKVTCLLYPHOVSARYDORYLISGVDPFVLIIDIPSGEMKHIFCVHGEITQLLVPP 522
DB 430 HEKEV-----NSVAFSPDQWIVSASNDSTIRLWDSNGNPIGQPMQGHEKEVNSVAFSP 483
QY 523 ENCSARVQHICSVASDHSVGLLSREKKCIMLASRHLFPQIVIKWRPSDDYLAVVCSDG 582
DB 484 DG-----QWIVSASNDSTIRLWDSNGNPTGQPMQGHEKEVNSVAFSPDQWIIISASND 537
QY 583 SVYVWQMDTGALDRVWG-----ITAVE-----ILNACDEAVPAAVDSLSPH 624
DB 538 TIRLWDSNGNPIGQPMQGHEKEVNSVAFSPDQWIIISASNDSTIRLWDSNGNP 590

RESULT 12

US-10-437-963-153705
; Sequence 153705, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 153705
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53635C.1.pep
US-10-437-963-153705

Query Match 2.6%; Score 204; DB 16; Length 891;
Best Local Similarity 19.6%; Pred. No. 2.5e-07;
Matches 161; Conservative 117; Mismatches 298; Indels 244; Gaps 36;

QY 22 ISAVLTDDGATVTCGHCQICLWDLVSELQINPALLFGHTASTTCL-----70
DB 80 VTALALSPDSRLLTAGHSRLIRVMDL-----ASRTCTRSWKGHDPGR 123
QY 71 SKACASSDKQIVTSASBSGEMCLWDVSDGRCIBFTKLACTHTGICQVQFVSGVGNRGRLL 130
DB 124 AMACHASG-GLLATAGADKKVCVMDVDDGFCFTHFR---GHAGV-----VTVVM 168
QY 131 CHGHPYPIILVV-----DATSLEVLYSLVSKISPDWISSMSIIRSHRQEDTVVALSVTG-- 184
DB 169 FHKDPKRLILFSGSEDAVTR--VWNLESK-----KCVAVLKEHFSAV-TSLALSEGGQT 219
QY 185 -----ILKWIWITSEISDMQDTEPIFEESKPIYQNCQSIISFCAPTQSRLLVVCVK 236
DB 220 LLSAGRDKIVNVDVRKYS--KKTIPAFE-----MIEDVSF-----254
QY 237 YMRVPDAGDYSLLCSGSPENGQWTGDFVSSDK---VLIWTEGQSYIYKLPASCLPAS 293
DB 255 -----IQPGSNLLSCLGEPANIKRKTGTYFLTVGERGVRWICLESQAQIYEQOS-----304
QY 294 DSFRSDVGKAVEN-----LIPVQHILLDRKXELLICPPVTRFFYGCRRYFKH 341
DB 305 -----SDVTYNTENESRRGFTSAVMSLDDQGLLCATADOQFLFYCSTRD-----350
QY 342 KLLIQDSSGRINLW-----NISDTADKQSBEGLAMTTSISLQEAFLDKLPCPAGI 393
DB 351 -----CGDF--QLNLYKRLVGYNDEILDLPFVGEDEQYLAVATNLEQVRVYDVA-- 398
QY 394 IDQLSVIPNSNPLKVTASVYIPAHGRVL-----CGREDGSIIVPATQTAIVOLLOGEH 448
DB 399 -----MSCSYVLSGHTEIIVVCIDTCISSSGKTLVWTGSKDSTVRL-----438
QY 449 MLRRGWPPHRTL-----RGHRNKVTCLLYPHOVSARYDORYLISGVDPFVSIW-----497
DB 439 -----WDMERRSCIGIGKHLGAISVAF-----SKSKNFFVSGSDRTIKLWSWDDTL 488
QY 498 -DIFSG---EMKHIFCVHGEITQLLVAPPENCARSVQHICSVASDHSVGLLSREKKCI 553
DB 489 DDVGSEVPLKAKAVVAHDKDINSLSVSPND-----GLVCSGSEDTACIWKLPNLVPS 542
QY 554 MLASRHLFPQIVIKWRPSDDYLAVVCSGSDSVVWQMDTGALDRVWGITA-----VILNA 609
DB 543 VVLKGHKGIVSVFSPVQCVITSSGDRTKVIWAVADGSCCLKTFEGHTSSVLRASFSLH 602
QY 610 CDEAVPAAVDSLSPHAPVNLKQAMTRRSALAKNMAHKLQTLA-----TNLLASEAD--- 662
DB 603 GTQFVSCGSDGL-----VKLWTIKTNECIATF--DKHDGKRWALAVGKKTENLATGCTDAVL 657
QY 663 -----KGNLPKYSHNSLMVQAIKTNLTDPDIHVLFFDFDEALIIQLITEASRPN 711
DB 658 NLWHDCTMEDKQEDFRKKEELLRGQLENAVSDVDY-----AKAIQ-LAFELRRPH 708
QY 712 TALISPENLQKASGSDKGGSFITG---KRAAVLFOQVKE 748
DB 709 RLLELFTQLCRESDLEDTIEKALIGLPKEGLRVLEYIRE 748

RESULT 13

US-10-369-493-19046
; Sequence 19046, Application US/10369493

; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19046
; LENGTH: 1136
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
US-10-369-493-19046

Query Match 2.6%; Score 202; DB 15; Length 1136;
Best Local Similarity 19.3%; Pred. No. 5.7e-07;
Matches 125; Conservative 83; Mismatches 250; Indels 190; Gaps 20;

QY 1 MAGNSLVLPVINGR-----KAPHCISAVLLTDDGATIVTGCHDQICLWDLSVEL 52
DB 580 IASGSLDKTIKLSRDRGLFRTLNHGHDAVYSFSPDGGQTIASGSGDKTIKLW----- 633

QY 53 QINPRALLFGHTASITCLSKACASSDKQYIVSASEGEMCLMDVSDGRCIEFTKLACTHT 112
DB 634 QTSQGVLLTKITGHEQTVNNVYSPDGKNLASASSDHSIKLWTTSG-----QLLWTLT 687

QY 113 GIOFYQFVGNQREGRLCHGHYPEILV-----DATSLVLYSLVSKISPDWTSMSIIR 168
DB 688 GHSAGVITVRSPDQGTIAAGSEDKTVKLMHRQDKLLKTLNG-----HQDWNSLSFSP 742

QY 169 SHRTQEDTVVALSVTLKWKIVTSEISMDQTEPIFEESKDIYQNCQISFCAFTQR 228
DB 743 DGR-----TLASADKTIKLMRIA-----DGKLV-----K 768

QY 229 SLVVCCKVWRFDAGYSLCLCGSPENG-----QWTGDFVSSDKVLIWTEN 277
DB 769 TLKHNDSVWDVNFSSDGAIASASRDNTIKLNRHGIELETFTG-----HS 815

QY 278 GQSY-YKLPASCLPASDSFRSDVGKAVENLIPVQHI----- 314
DB 816 GGVYAVNFLPDSNIIASASLDNTIRLQWRPLISPLEVLACNSGVYAVSFLHDGSIITAG 875

QY 315 -----LLDRKDKELLICPPVTFPFYCCREYFHKLIIQDSSGRIN-----IWNISDTA 362
DB 876 ADGNIQLMHSQDGSLLKTLFGNKAIYGI-----SFTPDGLIASANADKTVKLRVRD-- 928

QY 363 DKQSGEGLAMTTSISLQRAFDPKLNPCAGIIDQLSVIPNSNEPLKVTASVYIPAHGRLV 422
DB 929 -----GKALKTLIGHNEVKNVNSPDG-----KTLASASRDN-----TVKLNVSQGF- 973

QY 423 CGREDGSIIVPATQTAIVQLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARYDQ 482
DB 974 -----KKTLLKHTDEV-----FWVSFSPDG 993

QY 483 RYLISGVDPSVLIWDIFSGEMKHI PCVHGGEITQLLVPPENCARSVQHICISVADHSV 542
DB 994 KIITASADKTIIRLWDSFGSGLNLIKSLPAHNDLVYSVNFNPDG-----SMLASTSADKTV 1047

QY 543 GLLSLREKKCMLASRLHFPQIVIKWRPSDDLYLVGCSGSGSVYVWQMD 590
DB 1048 KLMRSHDGLLHFTSGHNSVYSSSPDGRYIASASEDKTVKIWIQID 1095

RESULT 14
US-10-369-493-11348
; Sequence 11348, Application US/10369493

; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11348
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
US-10-369-493-11348

Query Match 2.5%; Score 197.5; DB 15; Length 429;
Best Local Similarity 18.2%; Pred. No. 2.6e-07;
Matches 104; Conservative 85; Mismatches 172; Indels 209; Gaps 21;

QY 22 ISAVLLTDDGATIVTGCHDQICLWDLSVELQINPRALLFGHTASITCLSKACASSDKQY 81
DB 28 VFSIAVTPNGKAVSGSHDGTIAKVDLEKWRIRS---LRAHSKSIATAFA---ITSQGLK 81

QY 82 IVSASESGMCLMDVSDGRCIEFTKLACTHTGTFYQFVSGVGNQREGRLCHGHYPEILV 141
DB 82 VVLGSLDGNLWVNLTEGE-----EKAA-----FKEHS----- 109

QY 142 DATSLVLYSLVSKISPDWISMSIIRSHRTQEDTVVALSVTLKWKIVTSEISMDQOT 201
DB 110 -----EPITEIV---ITPD-----GKRAVSGSDNTLKW---DLEKMEEL 144

QY 202 EPIFEESKPIYQNCQISFCAFTQSLVVCCKVWRFDAGYSLCLCGSPENG-QTW 260
DB 145 TTLISHSN-----SVSKIAIT-----PSGKYAI---SGSDNTLKW 178

QY 261 TGQDFVSSDKVIWTENGQSYIYKLPASCLPASDSFRSDVGKAVENLIPVQHIILDRKD 320
DB 179 ---DLKCLDEEITSGHKSIV----- 196

QY 321 KELLICPPVTRPFYCCREYFHKLIIQDSSGRININWISDTADKQSGSEGLAMTTSISLQ 380
DB 197 NKIVITPD-----GKLAVSSSYDGTIAKVDLTKKEK----- 228

QY 381 EAFDKLNPACAGIIDQLSVIPNSNEPLKVTASVYIPAHGRVCGREDGSIIVPATQTAI 440
DB 229 -----VTLKGHSGP---VTDVFTPDGKRIISGSDKTLRVWD----- 263

QY 441 VQLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARYDQRYLISGVDPSVLIWDIF 500
DB 264 -----LKKG---NMTLKGKREVT-----SVAITSDGKYAISGSDFTIKVWDL 305

QY 501 SEMKHI PCVHGGEITQLLVPPENCARSVQHICISVADHSVGLLSLREKKCMLASRLH 560
DB 306 NGKIKVTLGHEKNYISTISIIPN-----KNCIVSSSHDETIAKVDLDRGIDITILIGH 359

QY 561 PFIQVWKWRPSDDLYLVGCSGSGSVYVWQMD 590
DB 360 GSVSSVAITPDGKSIIVSASGSGTHKIWSLE 389

RESULT 15
US-10-369-493-18905
; Sequence 18905, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:12:56 ; Search time 55 Seconds
(without alignments)
2606.600 Million cell updates/sec

Title: US-10-645-335-2
Perfect score: 7773
Sequence: 1 MAGNSLVPLVWGRKAPTH.....TSNRNVILMAHDGKEHRFMV 1490
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 791:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6023	77.5	1160	2 T00272	hypothetical prote
2	518.5	6.7	1446	2 T13018	hypothetical prote
3	234	3.0	1526	2 AC2239	WD-40 repeat prote
4	230.5	3.0	1227	2 AE1810	WD-40 repeat prote
5	229	2.9	1258	2 AI2155	WD-repeat protei
6	217.5	2.8	1101	2 T26919	hypothetical prote
7	202	2.6	1683	2 AF2071	WD-40 repeat prote
8	197.5	2.5	2241	2 T16064	hypothetical prote
9	192	2.5	1049	2 T42045	beta transducin-li
10	191	2.5	1151	2 T33777	hypothetical prote
11	189	2.4	494	2 T19550	hypothetical prote
12	188.5	2.4	1276	2 T18526	SREBP cleavage act
13	186	2.4	1356	2 T18521	beta transducin-li
14	182.5	2.3	1184	2 A96638	hypothetical prote
15	182	2.3	943	2 S59317	DIP2 protein - yea
16	181	2.3	1693	2 S76086	beta transducin-li
17	180.5	2.3	1747	2 AC1842	WD-40 repeat prote
18	180	2.3	1711	2 AD1842	WD-40 repeat prote
19	174.5	2.2	579	2 T22703	hypothetical prote
20	173.5	2.2	1216	2 H85023	hypothetical prote
21	172	2.2	981	2 T18234	beta transducin ho
22	170.5	2.2	876	2 T51507	WD40-repeat protei
23	170	2.2	1189	2 AI2493	WD-repeat protei
24	167.5	2.2	786	2 AG2375	WD-40 repeat-prote
25	166	2.1	2629	2 T30987	telomerase-associa
26	165.5	2.1	1223	2 T17345	hypothetical prote
27	165	2.1	677	2 AE1861	serine/threonine k
28	163.5	2.1	961	2 E86245	hypothetical prote
29	163	2.1	1194	2 T03818	apoptotic proteina

30	161.5	2.1	701	2 T16607	hypothetical prote
31	160	2.1	504	2 TS0983	probable pleiotrop
32	160	2.1	1189	2 AH2154	WD-repeat protein
33	156.5	2.0	530	2 T20360	hypothetical prote
34	156.5	2.0	571	2 T20359	hypothetical prote
35	153	2.0	765	2 T49346	conserved hypothet
36	152.5	2.0	1016	2 T19006	ankyrin related pr
37	152	2.0	788	2 S53923	probable membrane
38	151	1.9	515	2 S19487	hypothetical prote
39	151	1.9	1215	2 I52882	autoantigen - huma
40	150.5	1.9	499	2 A56021	probable cell divi
41	149.5	1.9	605	2 T38932	probable sulfur me
42	149	1.9	473	2 T33805	hypothetical prote
43	149	1.9	760	2 T41644	hypothetical trp-a
44	148	1.9	486	2 S49820	PRLL protein - Ara
45	146.5	1.9	589	2 AG2400	WD-repeat protein

ALIGNMENTS

RESULT 1

T00272
hypothetical protein KIAA0541 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C:Accession: T00272
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00272
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1160 <NAG>
A:Cross-references: UNIPROT:Q9Y4E6; EMBL:AB011113; NID:g3043605; PIDN:BA25467.1; PID:g
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0541

Query Match 77.5%; Score 6023; DB 2; Length 1160;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	331	REFYGCREFYFKLLIQGSSGRINIWNISDTADKQSGSEGLAMTTSISLQEA	FDKLNPCP	390
DB	1	REFYGCREFYFKLLIQGSSGRINIWNISDTADKQSGSEGLAMTTSISLQEA	FDKLNPCP	60
QY	391	AGIIDQLSVIPNSNEPLKVTSVYIPAHGRLVCGREDGSIIVPATQTATVQLLQ	GEHML	450
DB	61	AGIIDQLSVIPNSNEPLKVTSVYIPAHGRLVCGREDGSIIVPATQTATVQLLQ	GEHML	120
QY	451	RGWPPHRTLGRHNKVTCLLYPHQVSARYDORYLISGGVDFSVIITWDFSG	BMKHFCV	510
DB	121	RGWPPHRTLGRHNKVTCLLYPHQVSARYDORYLISGGVDFSVIITWDFSG	BMKHFCV	180
QY	511	HGGEITQLLVPPENCARSARVQHICICSVASDHSVGLSLREKCKICMLASR	HLFPQIKWRP	570
DB	181	HGGEITQLLVPPENCARSARVQHICICSVASDHSVGLSLREKCKICMLASR	HLFPQIKWRP	240
QY	571	SDYLVVSGSDGSVYVWMDTGALDRCVNGITAVEILNACDEAVPAVDSLSH	PAVNLKQ	630
DB	241	SDYLVVSGSDGSVYVWMDTGALDRCVNGITAVEILNACDEAVPAVDSLSH	PAVNLKQ	300
QY	631	ANTRSLAALKMAHHKLTATNLLASEADKGNLPKYSNLSMWQAATKTNLTDP	DIHV	690
DB	301	ANTRSLAALKMAHHKLTATNLLASEADKGNLPKYSNLSMWQAATKTNLTDP	DIHV	360
QY	691	LFDFVEALIIQLITEASRPNTALISPENLQKAGSSDKGGSFLTGKRAAVLF	QQVKETI	750
DB	361	LFDFVEALIIQLITEASRPNTALISPENLQKAGSSDKGGSFLTGKRAAVLF	QQVKETI	420
QY	751	KENIKEHLDDDEDEBEINRQRRSSDPYRSSKSKPLTLLEYNLTMDTAKFL	MSCLHAW	810

Db	421	KENIKHELLDDEEBEIMRQREESDPYRSKSKPLTLEYNLTWDTAKLFMSCLHAW	480
Qy	811	GLNEVLDEVCLDRGLMLKPHCTVTSFGLSRGGHMSLMLPGYNOPACKLSHGKTEVKRLP	870
Db	481	GLNEVLDEVCLDRGLMLKPHCTVTSFGLSRGGHMSLMLPGYNOPACKLSHGKTEVKRLP	540
Qy	871	ASGVGKGTVGSRAVTTQHLLSIIISLANITMSTNATFTIGDHMKKGTPRPSPSTDLS	930
Db	541	ASGVGKGTVGSRAVTTQHLLSIIISLANITMSTNATFTIGDHMKKGTPRPSPSTDLS	600
Qy	931	KARGSPPTSSNIIVOGQIKQVAAPVVSARSADSHSGSDPPSAPALHTCFLVNEGWSQLAAM	990
Db	601	KARGSPPTSSNIIVOGQIKQVAAPVVSARSADSHSGSDPPSAPALHTCFLVNEGWSQLAAM	660
Qy	991	HCVMPLDLLGLDKFRPPLLEMLARRWDRCLEVREAQAALLLAELRIIIOAGKEAIDAW	1050
Db	661	HCVMPLDLLGLDKFRPPLLEMLARRWDRCLEVREAQAALLLAELRIIIOAGKEAIDAW	720
Qy	1051	APYLPOVIDHVISPGVTSEAAQTITTAPDASGEAKVQEEHDLVDDDIITTCGLSSVPQM	1110
Db	721	APYLPOVIDHVISPGVTSEAAQTITTAPDASGEAKVQEEHDLVDDDIITTCGLSSVPQM	780
Qy	1111	KKISTSYEERRKQATAIVLLGVIGAEFGAIEPPKLLTRPRSSQIPGEGFLTSGGSNY	1170
Db	781	KKISTSYEERRKQATAIVLLGVIGAEFGAIEPPKLLTRPRSSQIPGEGFLTSGGSNY	840
Qy	1171	LARHTKALTFLLOPPSPKLPHSTIRRTAIDLIGRGFTTWEPYMDVSAVLMGLELCA	1230
Db	841	LARHTKALTFLLOPPSPKLPHSTIRRTAIDLIGRGFTTWEPYMDVSAVLMGLELCA	900
Qy	1231	DAEKOLANITWGMLPSPADSARSARHALSLIATARPPAFITTIKEVHRHTALAANTOS	1290
Db	901	DAEKOLANITWGMLPSPADSARSARHALSLIATARPPAFITTIKEVHRHTALAANTOS	960
Qy	1291	QQNHMTTLARAKEILRVIEILLIEXMPTDVVDLLVEVMDIIMYCLESGLVKKKGLOECF	1350
Db	961	QQNHMTTLARAKEILRVIEILLIEXMPTDVVDLLVEVMDIIMYCLESGLVKKKGLOECF	1020
Qy	1351	PAICRFPMWSYYERNHRIAIGABHGSVALYDIRTGKQOTIHGHKGPITAVAFAPDGRYLA	1410
Db	1021	PAICRFPMWSYYERNHRIAIGABHGSVALYDIRTGKQOTIHGHKGPITAVAFAPDGRYLA	1080
Qy	1411	TYSNTDSHISFWQNTSLGSIWMNSAPOLRCIKTYQVPPVQPASPFSHNALKLARLIW	1470
Db	1081	TYSNTDSHISFWQNTSLGSIWMNSAPOLRCIKTYQVPPVQPASPFSHNALKLARLIW	1140
Qy	1471	TSNRNVILMAHDGKEHRFMV 1490	
Db	1141	TSNRNVILMAHDGKEHRFMV 1160	

RESULT 2
 Tl3018
 hypothetical protein F8L21.60 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
 R;Accession: Tl3018
 R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirks, W.; Stiekema, W.; Bancroft, I.; Mey
 submitted to the Protein Sequence Database, July 1999
 A;Reference number: T17587
 A;Accession: Tl3018
 A;Molecule type: DNA
 A;Residues: 1-1446 <BEV>
 A;Cross-references: UNIPROT:Q9SU74; EMBL:AL096982; GSPDB:GN00062; ATSP:F8L21.60
 A;Experimental source: cultivar Columbia; BAC clone F8L21
 C;Genetics:
 A;Gene: ATSP:F8L21.60
 A;Map position: 4
 A;Introns: 895/3; 930/2; 953/3; 1221/3; 1289/2; 1330/3; 1399/3

Query Match	6.7%	Score 518.5;	DB 2;	Length 1446;
Best Local Similarity	19.9%;	Pred. No. 1.3e-27;		

Query Match 6.7%; Score 518.5; DB 2; Length 1446;
Best Local Similarity 19.9%; Pred. No. 1.3e-27;

QY 912 DHMKGTRPRRSTPDLKARGSPPTSSNIVGQIKQVAAVVSARSADHSGSDPPSA 971
|||:
Db 926 -----ATYSSNLA----- 933
QY 972 PALHTCFVNEGWSQLAAHCVMLPDLGLDKRPPLEMLARRWDRCLVEVEAAQALL 1031
|||:
Db 934 -AYYTRNLAEK-----YDDL-----KPLLQLLVTFWQDSEQVRAARSUF 974
QY 1032 --LAELRRIRQAGRKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITTPADSGPEAKVQE 1089
Db 975 HHTASL-----AIPPLCSDH-----ASEHAELVRSLSGISLNEPKVLS 1013
QY 1090 EHDLDVDDITTTGCLSS--VPQMKKISTSYEERRKQATAILVGLVGAEBGABIEPPKLL 1147
|||:
Db 1014 ----TGEYPTNSLSEHIOARLSQABE-----SEILSWLE----- 1047
QY 1148 TRPSSSOIPEGFCLTSGGNSYLSARHTCKALTF-----LLLOPPSPKLPPTHSTIR----- 1198
Db 1048 ----SPEMDWTISCVCVGTSDADAAHIIIVAAALSIIWYPSLVKPGMLVVKLLNLVMA 1102
QY 1199 -----RTAIDLIGRGP-TWEPYM--DVSALVLMGLLELCADAEKQLANITMGL--PLS 1246
|||:
Db 1103 MSEKYSTAAEALLSEGMEITTKTWIGDPDIPRV-----SDIFFQIECVSSSVGAYQVV 1155
QY 1247 PAADSARSARHALSIATARPPAFITTIKVEVHRHTALAAANTOSQONMHTTILARAKEI 1306
|||:
Db 1156 PSSIKETVELVPLSLAMADVLGFLSIIESQIW-----STASDSPVHVVSILR-----TL 1204
QY 1307 LRVE-----TLIEKMPDVVDLLVEVD-----IIMYCLEGSLVKKGLQCECPAIC 1354
|||:
Db 1205 IRIIRAAPRNVLHLEK-----AVNFVLQTMPSNTVMKTKLQTSMATLREVVRVFP----- 1257
QY 1355 RFYMWVSYERNHRIAVG-----ARHGSVALYDRTGKQTIHGHKGP----- 1396
Db 1258 ----MVLNDSSTRLAIGDVTETNNACIHIYDNRMTKIRVLDASGPGLPNPLRGASES 1314
QY 1397 ----ITAVAPADPGRYLATYSNTDHSIFWQMTSLGSI-----GMLNSAPQLRCIKTY- 1447
|||:
Db 1315 AVTTAISALSFSPDGEGLVAFSENGLMRWWS-----LGSVWWEKLSQSITPQCTKLIF 1369
QY 1448 -----QVPPVQAPSGHNSALKIARLI-----WTSN 1473
Db 1370 IHPWDGFSSSSRSTSVISSINDEQELPQETAKNISHVERLQVLQVHLDLSYRLEWASE 1429
QY 1474 RNVILMAHD 1482
Db 1430 RKVVLTRHN 1438

RESULT 3

AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AC2239
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2239
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <KUR>
A;Cross-references: UNIPROT:Q8YR11; GB:BA000019; PIDN:BAB75165.1; PID:g17132599; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3466

Query Match 3.0%; Score 234; DB 2; Length 1526;
Best Local Similarity 20.1%; Pred. No. 1.8e-07;

Matches 132; Conservative 100; Mismatch 228; Indels 196; Gaps 29;
QY 19 THCISAVLLTDDGATIVTGCDDQICLWDLSELQINPRALLFGHTASITCLSKACASSD 78
|||:
Db 1032 TSCVRSVVFSSDGMALSGDDQTVRLMDIS---SGNCLYTLQGHT---SCVRSVVFSPD 1085
QY 79 KQYIVSASSGEMCLWDVSDGRCIEFTKLACTHTGTQYQYFSGVGNOREGRLCHGYPEI 138
|||:
Db 1086 GAMLASGGDQIVRLMDISSGNCL-YTLOQYT-SWRFVLES---PNGVTLANGSSDQI 1139
QY 139 L-VVDATSEVLSYLSVSKISPDWISSMIIIRHRTQEDTVWALSVTG---ILKVVIVTSEI 195
|||:
Db 1140 VRLNDISSKCLYTLQG--HTWVNAVAF-----SPDATLASGSDQTVRLMDISS-- 1189
QY 196 SDMQDTEPIFEBSKPIYCNQCOSISFCAPTQSLVIVCSKYW---RVFDAGDYSLSCSG 252
|||:
Db 1190 ----SKCLY-----ILQHTSWVNSVVFNP-DGSTLASG 1218
QY 253 PSENGQTWTGGDFVSSDKVI-TWENGQSYIYKLPASCLPASDSFSDVGVKAVENLPPV 311
|||:
Db 1219 ----SSDQTVRLWEINS-----SKCLCTFQGHTSWNSVVFN----- 1251
QY 312 QHILLDRKDELLICPPVTRFFVGCREYFHKLLIQDSSGRNLNINWISDTADKQSEGL 371
|||:
Db 1252 ----PD-----GSMLASGSDKTVRLMDISSKCLHTFQGHT 1284
QY 372 AMTTSISLQEAFFKLNMP---CPAGIIDQ---LSVIPNS-----NEELKVYASYIIPA 417
|||:
Db 1285 NVNNSVA-----FNPDSGMLASGSDQTVRLWEISSKCLHTFQGHTSWNSVVFSPD 1337
QY 418 HGRVLCGRDGSIVIVPATQTAIVQLQGEHMLRQWPPHRTLRGRNKVTKCLLYPHQVVS 477
|||:
Db 1338 GTMLASGSDQTV-----RLWSISSGECLY-----TFLGHTNMGVSVIFSP--- 1378
QY 478 ARYDQRYLSGGVDFSVIIVDFISGEMKHFVCHGGEITQLLVPPNCSARVQHCICVSA 537
|||:
Db 1379 ----DGAILASGSDQTVRLWSISSGKCLYTLQGNHNSVSVFSPDGT-----LIASGS 1429
QY 538 SDHSVGLSLRKKCKIMLASRHLFPQIVIKWRPDDYLVVCGSDGSVYVQMMDTGALDRC 597
|||:
Db 1430 DQOTVRLWNISGECLYTLGHINSVRSVAPSSDGLILASGSDDETIKLWDVKTG---EC 1486
QY 598 V-----MGITAVELINACDEAVPAVDLSLHAPVNLKQAMTRSLAALKN 642
|||:
Db 1487 IKTLRSEKIVEGNITSVRGLTEVERA-----TLKTLGAVEN 1523

RESULT 4

AE1810
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AE1810
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE1810
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1227 <KUR>
A;Cross-references: UNIPROT:Q8ZOR1; GB:BA000019; PIDN:BAB77553.1; PID:g17135007; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr0029

Query Match 3.0%; Score 230.5; DB 2; Length 1227;
Best Local Similarity 19.6%; Pred. No. 2.2e-07;
Matches 142; Conservative 102; Mismatch 248; Indels 231; Gaps 31;

QY 22 ISAVLLTDDGATIVTGCDDQICLWDLSELQINPRALLFGHTASITCLSKACASSDKQY 81

Db 605 VSVKSPDKGYATGLMGEIRLWQTSNKKQLR---IYKGTAWWAF---FSPDSRM 658
QY 82 IVSASEGEMCLWDVSDGRCIEFTKLAHTGTGTFYQFVSGNOREGLLCHGYPE-ILV 140
Db 659 LASGSADSTIKLMDVHTGECL---KTLSKNTN---KVYSVAFSPDGRILASASQDTIKL 712
QY 141 VDATSLEVLVSKISPDWISSMSIIRSHRTQEDTVVSLVSGILKLVIV-----191
Db 713 WDIATNGCOQTLLG---HDDWVSVTFSPVDDRLPLLLASSADQHIKJMDVATGKCLKTL 770
QY 192 ---TSET-----SDMDTE-----PIFEESKPIY---CONCOS 219
Db 771 KHTREHVSVPSPDGGTGLASSGEDSTVRLWDVKTGCQWQIFGSHKKYIVRVFSPDGT 830
QY 220 ISPCAFQTRSLV-----VC-----SKYWRV-FDAGDYSLLCSGSPSENGQTTWGD 264
Db 831 LASCG-EDRSIKLWDIQRCGCNVTLWGHSSQVWAIAFSPDGRTLISCSDDQTARLW---D 886
QY 265 FVSDDKVII---WTENGQSIIYKLPASCLPASDSFRSDVGVKAVENLIPVQHILLDRKDK 321
Db 887 VITGNSLNLIRGYTRDYSVAFS-PDSQILASG--RDDYTIGLWNLKTGECHPLRGHQR 943
QY 322 ELLICPPVTRFFYGCREYFH---KLLIQGDSGRLNINISDTADKQSGEGLAMTTSIS 378
Db 944 IRSVA-----FHPDGKILASGADNTIKLWDIST-----973
QY 379 LQAFDKLPCPAGIIDQLSVIPNSNEPLK-----VTASVYIIPAHGRLVCGREDGSI 430
Db 974 -----NHSKYIRTLGTHTNWMVTVFSPDKHTLASSSEDRTI 1010
QY 431 VIYPATQTAIVQLLOGEHLMLRCWPPH-----RTLGRHKNKVTCLLYPHQVSAARYDYL 485
Db 1011 -----RLWDXDTGDCLOKJGSHWVWTVAFSP-----DGRIL 1043
QY 486 ISGVDPSVLIIDIFSGEMKHIFCVHGEITQLLVPPENCARSVQH-----CICSVASDH 540
Db 1044 ASSADSEIKLMDVASGK-----CL-----QTLTDPQMIWSVAFSLDGTLLASASEQ 1092
QY 541 SVGLLSRLREKKICIMLASRHLFPFQIVIKWRPSDYLIVVCGSDGSVYVQMDGTALDRCVMG 600
Db 1093 TVKMLNLTGECVHTLKGHEKQVYSVAFSPNGQIAASGSEDTTVKLDWIDISTG-----1144
QY 601 ITAVEILNACDEAVPAVDLSHPAVNLKQAMTRRSIALKNAHKLQTLATNILLASEA 660
Db 1145 -----SCVDTLKH-----GHTAAIRSVAF-----SPDGRLLASGS 1174
QY 661 SDK 663
Db 1175 EDE 1177

RESULT 5

AI2155
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AI2155
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2155
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1258 <UNP>
A;Cross-references: UNIPROT:Q8VTC2; GB:BA000019; PIDN:BAB74499.1; PID:g17131893; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2800

RESULT 6

T26919
hypothetical protein Y45F10B.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T26919
R;McMurray, A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z20286
A;Accession: T26919
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1101 <WIL>
A;Cross-references: UNIPROT:O62471; EMBL:AL021487; PIDN:CAAL16357.1; GSPDB:GN00022; CESP
A;Experimental source: clone Y45F10B
C;Genetics:
A;Gene: CBSP:Y45F10B.10
A;Map position: 4
A;Introns: 217/3; 292/1; 451/3; 556/2; 637/3; 755/3; 823/2; 946/1; 1078/3
Query Match 2.8%; Score 217.5; DB 2; Length 1101;
Best Local Similarity 18.0%; Pred. No. 1.5e-06;
Matches 149; Conservative 110; Mismatches 246; Indels 325; Gaps 35;

QY 16 KAPTHCISAVLLTDDGATIVTGCCHDQICLWDLSEVQLINPRALLFGHTASITCLSKACA 75
DB 417 KGHTAAVYVTSCLSSNDLSFLVSTSPDKTVNVWVFS---QSTPTMSLTHHTAKVTC---AIL 470
QY 76 SSDKQYIVSAS-----ESGEMC-----L 93
DB 471 TSDQYILITASADSSAKMIKLEGEVMSFNDHTGVSVLSQLTNNQFLITGSGDFVQM 530
QY 94 WDVSQDGRGIE-----FTKLACTHTG:QFYQFVSQVGNQREGRLLCHGH 134
DB 531 WDVNTGKICRGMGLMAPVSTLAITSNDFAFWVACEDETLKVSF-TVGGQELHELMHEG 589
QY 135 YPELLVVDATSLVLYLSVSKI-----SPDWISSMSIIRSHRTQEDTV 177
DB 590 KWSLVCAQDDCOLFAATKSKVFCYDIHNGQIMIDLPTAQFPFICSLKI-----SSDNY 643
QY 178 VALSVTC-ILKMWIVTSEISDMOD-----TEPIFEESKPIYC-----Q 215
DB 644 FLISPCGPKVTIWNVTKRNDHADVHADKEGFLTAVALSNDKDYAACGTNGGIVALWDL 703
QY 216 NCQISISFCAPT---QRSLVVCVKYRVFDAGDYSLLCSG-----PSENG-----257
DB 704 VCO---CVPTIQNKGDPTICRY-----SVDSQYICISGNOAGCILLILDAQNGGVREL 754
QY 258 -----QWTGDFVSSD---KVIWTEGQSVIYKLPASCLPASDSFRSDVGKAV 304
DB 755 FMHSSEVLSMSLVHNMKMSCDIQGMVWELFGDDPTPEMVAT-----798
QY 305 ENLIPVVOHILLDRKOKELICPPVTRFFYGCREFYHKLIIQDSSGRLNIWNISDTADK 364
DB 799 -GVKPP1-----FVPTGRIMVG-----HCSL-----SNKEMKIWF-----829
QY 365 QGSEEGLAMTTSISLOBAFDPKPCPAGIIDLQSVIPNSNEPLKVTASVYIPAHGRILVC- 423
DB 830 --PEEGPPVTRA-----KLS-----HSDS-----ITCFATSPKGFNIAT 862
QY 424 GREDGSIIVPATQTAIVQLQGHMLRGRWPHRTLRGRNKVTCCLLYPHQVSARYDQ 483
DB 863 GSRDMSLKIWQIDKGFLTVLV-----GHENVVTC-----CCISPFDER 900
QY 484 YLISGGVDFSVIIVDFSGEMKHFVCHGGEITQLVPPENC SARVQHOCICSVASDHSVG 543
DB 901 LVVSGARDEKIIIVNVQSGDM-----VCTVNTTAAT 932
QY 544 LLSLRKKKCIIMASRLHFPPIQVIKWRPDDYLVVCGSDGSVYVQMDTGAL-----594
DB 933 SLSTGDSVTVFST-----TEDGWVETWSTTKGRLSTFNARHP 971
QY 595 -DRVCMGITAVEILNACDEAVPAVDSLSH-PAVNLQAMTRRSALAKNMAHKKQLTLA 652
DB 972 IKKLINSYSHRMLLLLENCAQLPILCLHNTPAVGVE--ATRRRSARAQSVSSASNEPVA 1029
QY 653 TN-----LLAS-----EASDGNLPKYSHNSLMVQAIKNTLDDP 687
DB 1030 STSAGEIKKDPILLSSNNGNAQAPRATAPKPTFDMLERSKSRSLIEKD 1079

RESULT 7

AF2071
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AF2071
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1683 <KUR>

A:Cross-references: UNIPROT:Q8YV57; GB:BA000019; PIDN:BA073823.1; PID:g17131215; GSPDB:
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all12124

Query Match 2.6%; Score 202; DB 2; Length 1683;
Best Local Similarity 19.3%; Pred. No. 3.8e-05;
Matches 125; Conservative 83; Mismatches 250; Indels 190; Gaps 20;
QY 1 MAGNSLVLPVLMGR-----KAPTHCISAVLLTDDGATIVTGCCHDQICLWDLSEVQL 52
DB 1087 IASGSLDKTIKLSRDRGLFRTLNGHEDAVYVSFSPDGTQTIASGSDKTIKLM-----1140
QY 53 QINPRALLFGHTASITCLSKACASSDKQYIVSASEGEMCLWDSQRCIEFTKGLACTHT 112
DB 1141 QTSDDGTLTKITIGHEQTVANNVSPDPGKNLASASSDHSIKLMDWTSG-----QLLMTLT 1194
QY 113 GIQYQFVSQVGNQREGRLLCHGHYPEILV-----DATSLVLYLSVSKISPDWISSMSIIR 168
DB 1195 GHSAGVITVRFSFDPGTIAAGSEDKTVKLRHQRQDKLLKLTNG-----HQDWNSLSFSP 1249
QY 169 SHRTQEDTVVALSVTGILKMWIVTSEISDMQDTEPIFEESKPIYCQNCQISFCAPTQ 228
DB 1250 DGK---TLASASADKTIKLMRIA-----DGKLV-----K 1275
QY 229 SLVVCVKYRVFDAGDYSLLCSGSENG-----QWTGDFVSSDKVLIWNTEN 277
DB 1276 TLKGHNSVMDVNFSSDGGKAIASASRDNTIKLWNRHGLELETFGT-----HS 1322
QY 278 GQSY-IYKLPASCLPASDSFRSDVGKAVENLIPVQHI-----314
DB 1323 GGYAVNFLPDSNIIASASLDNTIRLQRPISLPLEVLGNSGVYVSLFHGDSIIATAG 1382
QY 315 -----LLDRKOKELICPPVTRFFYGCREFYHKLIIQDSSGRLN-----IWNISDTA 362
DB 1383 ADGNIQWHSQDGLSLKTLPGNKAIYGI-----SFTQGDLIASANADKTVKLWRVD-- 1435
QY 363 DKQSEEGLAMTTSISLOBAFDPKPCPAGIIDLQSVIPNSNEPLKVTASVYIPAHGRLV 422
DB 1436 -----GKALKTLIGHDNENKVNFPDQ--KTLASASRDN-----TVKLMVNSDGKF- 1480
QY 423 CGREDGSIIVPATQTAIVQLQGHMLRGRWPHRTLRGRNKVTCCLLYPHQVSARYDQ 482
DB 1481 -----KTLKGHTDEV-----FWVSFSPDG 1500
QY 483 RYLISGGVDFSVIIVDFSGEMKHFVCHGGEITQLVPPENC SARVQHOCICSVASDHSV 542
DB 1501 KIITASADKTIKLRWDSFGNLIKSLPAHNDLVYSVNFPDQ-----SWLASTSADKTV 1554
QY 543 GLLSLRKKKCIIMASRLHFPPIQVIKWRPDDYLVVCGSDGSVYVQMD 590
DB 1555 KLRSHDGHLLHTFSGHSNVVYSSSPDPGRIVASASEDKTVKIWD 1602

RESULT 8

T16064
hypothetical protein F13H8.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16064
R:Ding, H.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid F13H8.
A:Reference number: Z18455
A:Accession: T16064
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2241 <DIN>
A:Cross-references: EMBL:U23139; NID:g722370; PID:g722372; PIDN:AAC46678.1; CESP:F13H8.
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F13H8.2
A:Introns: 52/3; 178/2; 269/3; 336/2; 486/3; 701/3; 862/2; 938/3; 1033/1; 1074/2; 1117/3

1; 2185/2

Query Match 2.5%; Score 197.5; DB 2; Length 2241;
Best Local Similarity 17.9%; Pred. No. 0.00013;
Matches 168; Conservative 121; Mismatches 345; Indels 303; Gaps 36;

QY 9 PIVLWGRKAPTHCISAVLLTDDGATIVTGCHDGOICLWDLSELQINPRAL--LFQHTAS 66
DB 111 PIMFSGHKKAVNIE---FSSDGLLPATGKGQGVIVLWDLAE---RGMFLHGHKES 162
QY 67 ITCLSKACASSDKQYIVSASEGEMCLWVSDGRCI-----EFTKLAC 109
DB 163 VTQMKFV---RGDRFIISTKSLIKFWSIASQSCFVTWMSRNEIYSMSLHRDETFLVA 219
QY 110 THGTIQYQFSVG----- 122
DB 220 ATAELLLVLFELNWKNGAIFEQEVKSEPSAKKATNDEKEDDLANMANRYTTTKLRGR 279
QY 123 --NOREGRL--LCHGHYPELIV--VDATSLVLSVSK----- 155
DB 280 VIRQSKGRALQVATHDDRLVCVGADKQVADVTVFSETESAKELTKLKSARKKATSDT 339
QY 156 -----ISPDWISSMSIIRSHRTQEDTVV-----ALSVTGILKW 189
DB 340 TTVNEDDVAKDVTILVTRIGEMVPHKIKWIDVRSVKTNENTLIYKWTAL-LTDNTVHW 398
QY 190 IVTSEISDMQDTEPIFEESKPIYQNCOSISFCAPTQ--RSLVVCVKYWRVFDAGDYS 247
DB 399 VKMS-----IETTSNSVECDSVGNDLKLGHREDVRLACVSSSS-----S 437
QY 248 LLCGSPSENGQWTGDFVSSDKVITWTENGQSVYIKLPASCLPASDSFRSDVYKAVENL 307
DB 438 LLASG-----GG-----NEVIVNTHSL-----RSSLTLDNDIKDI--VAVNF 474
QY 308 IPPVOHILLDRKDELLICPPVTRFFYGCREYFHKLLIQDSSGRNLNINISPTADKQGS 367
DB 475 VPGDNYILTTGGKNGEI-----GVPELSSAELVETKKAHTCAITWIIQNSPNEG- 522
QY 368 BEGLAMTTSISLOEAFDKNPCPAGIIDQLSVTPNSNEPLKVTVASVYIPAHGELVCGRED 427
DB 523 ----FITASADKTVRFWSFVLVTEGSRKRISI-----KEQKVLELPDEALAAFSAD 570
QY 428 GSIVIVP-ATQTAIVOLLQGEHMLRRGPPHRTL-----NHLKPFVSLYGHSLPVC-----VDIAP 611
DB 571 GKFLVALLNNTCSVPV----- 611
QY 481 DORYLISGGVDFSVIIVDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHCICSVASDH 540
DB 612 SSKLCVTGSDVSKVWGLDFGCHKSFHAHDDAVTSVLFPCPE-----EQLFWSAGKDG 666
QY 541 SVGLLSLREKCCIMLASRHLFPPIQVWKRPSSDDYLVVVGSDGSVYVWQMDTGAL----- 594
DB 667 KIKQMDAVKFIIMQVLDRLHTLDIRCLAQFTNGSVMFSSASHDKSIRCWEKTDLEILVEERE 726
QY 595 -----DRCVMGITAVEILNACDEAVPAVDLSLHPAVNLQKQATRSLSAALKNMAH 645
DB 727 EVREBEYEKKLIEEDVVAAGENAE-----GVASKTSVVVSASNI- 772
QY 646 HKLQTLATNLLASEADKGNLP-----KYSHNSL-----MVOAI-KTNLTDPDIHVLFPDVE 696
DB 773 -QAVDIARNEKVKQEDPGHEPHEPLIGAYNSKSLDHPIDVVAACRASLDRLTLVPLS 831
QY 697 -----ALIITQLTEASRPN-----TALISPENLQKASGSSDKGGSFLTGKRAAVLFOQV 746
DB 832 YVADIILATISSCTQQQKAECLCTHVAVYLTRIHLSHITASSD-----YVPIFEQM 881
QY 747 KETIKENIK-----HLLDDEEEDIEIMRQ 771
DB 882 KSKMSGEVKLRNVGTGILNLAALRILLATELEDEQVKK 918

RESULT 9
T42045

beta transducin-like protein homolog - Streptomyces coelicolor
N:Alternate names: WD-40 repeat protein
C:Species: Streptomyces coelicolor

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T42045

R;Vachalova, K.; Stoicheva, Z.; Tichy, P.; Petricek, M.

A:Description: Cloning and characterization of the WD-40 repeat protein-encoding gene (1

A;Reference number: Z22038

A;Accession: T42045

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1049 <VAC>

A;Cross-references: UNIPROT:Q9ZEM4; EMBL:AJ131817; PIDN:CAA10512.1

C;Genetics:

A;Note: wdpA

Query Match 2.5%; Score 192; DB 2; Length 1049;

Best Local Similarity 20.5%; Pred. No. 8.7e-05;

Matches 137; Conservative 74; Mismatches 208; Indels 250; Gaps 33;

RESULT 10

T33777

hypothetical protein H24G06.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33777
R;Ryan, E.; Delehaunty, A.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid H24G06.
A;Reference number: 221406
A;Accession: T33777
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1151 <RYA>
A;Cross-references: EMBL:AF101309; PIDN:AA69213.1; GSPDB:GN00023; CESP:H24G06.1
A;Experimental source: strain Bristol N2; clone H24G06
C;Genetics:
A;Map position: 5
A;Introns: 18/3; 91/3; 220/3; 353/2; 441/3; 470/1; 510/2; 593/3; 773/2; 825/1; 852/2; 87

Query Match 2.5%; Score 191; DB 2; Length 1151;
Best Local Similarity 18.4%; Pred. No. 0.00012;
Matches 218; Conservative 163; Mismatches 376; Indels 426; Gaps 59;

QY 15 RKAPTHCISAVL---TDDGATVTC-HDQQLCLW-----DLSVELOINP-RALLFGH 63
DB 129 RSMQSHVTSITIVRPTDSTLISGQHDQAQIFTDWRTGHDLGKLMSPVNAITISF 188

QY 64 TASITCL-----SKACASDQYI--VSAS--ESGEMCL 93
DB 189 DSSM-CVTGSKSVKYWFLASDGMKRTGLTSRAILLADKRNVNFDVAACDSTNRW 247

QY 94 WDVSDGRCTEFT---KLACTHTGIGVQPSV-----GNQREGLLCHGHYPRILVVD 142
DB 248 TVTSSGEILFNGKVKVCYV---QFNFFVQVVKYTWKESGDFRAICIAIARVPEGLVG 304

QY 143 ATSLVLYSLVSKISPDWISSMIIIRSHRQEDTVVALSVTGLKWIYVSE---ISDM 198
DB 305 CSD-----GLVRFLCLIGDLDLADL 326

QY 199 QUTPEIFEEBSPKYQNCQSIFCAFTQRLSLVVCVKYWRVFDAG--DYSLLCSGSPEN 256
DB 327 APPTHLFQDPSNVVDYDQQLNHPEDALFPEPRCLVASSSSPTFVGVADRSLI----- 379

QY 257 GQTWTGDFVSSDKVIWENGQSYIKLPASCL---PASDSFRSDVGKAVENLIPVQH 313
DB 380 -----EFARENK--SWSPRASLGHGTGSVNCIEPPSPSSP----- 412

QY 314 ILLDRKDKELLICPPVTRFPYGCREFHKLIIQDSSGLNINIS-DTADKQSGEGLA 372
DB 413 -----CLPI-----GTLITGGSDGTIRFWNFGDSEDKK--BIANI 446

QY 373 MTTTSIQEAFDKLPCPAGIIDLQ---SVIPNSNEPLKVTAS---VYIPAGR-LVCG 424
DB 447 LCPSL-LKVTF--LDENPDLLVDKRNLESVGPENQSSGVLCTRVSHDGRMLIAG 503

QY 425 REDGSIVIPA--TOTALVOLLOGE----- 449
DB 504 TATGMLYVDLSFSDPIIDVINAHBDVTSIDFSDBASTSSHDHPIFLASGGRDRFVH 563

QY 450 LRKGWPPHRT-----LRGRNKVTCLLYPHOVSARYDORYLISGVDFSVIWDIFS- 501
DB 564 FRR--IPYSSQFVHCVAVLGDHQAISIKF-----ASNNGQLHLTYAASDRSLIWKLSF 617

QY 502 -----GEM--KHI--FCV----- 510
DB 618 SDQCHETFRVQMLSVASSIGDMNFIKHVDLFFVAGHDMRLRQPDINGKTVREKGTDDVD 677

QY 511 -HGBEITQLLAVPPENCARSARVQHICSVASDHSVGLLSLRKCKICMLASRLHLPFIQVKNR 569
DB 678 AHSKILKVAIDHSGSA-----ISVCSDKFVVVTDLRSGVCLAVLCGFGAPDATP- 730

QY 570 PSDDY--LVGCSGSGSVYVQMDTGALDRCMG-----ITAVEILNACDEAVP 615
DB 731 -SDDFKNVITVTSNGSIFIWQLAKNLTERMISAQVRLMEVTRTATPDPSLLSGSGSETIS 789

QY 616 AAVDSLSHPAVNLKQAMTTRSALAAKNMAHKLQTLATNLNLASEASDKGNLPKYSNLSM 675
DB 790 GDSNSFGRP-----LCAPD-----FSGSSASLYSDDDDDSTRFSS--- 824

QY 676 VQAIKTNLTDPDIHVLFPDVE-----ALIIQLLTEESRPNTALISPENLQKASGS 726
DB 825 VRSSRTKRILPNGHGLIGDSSYARVGDSSFSFAVOSAPAVERTHTNLFSHD--QYETDV 882

QY 727 SDKGSFLTGKRAAVLFOQVKETIKENIKELHLLDDEEDEI-----MKRREESD 777
DB 883 SETQSDFFVSSR-----KTRLPSEDDQDNLGSAQYLAAPINEDRRSAS 926

QY 778 PE-----YRSKSKPLTLLEY---NLTWDTAK-LFMSCLHAW-----GLNEVLDE 818
DB 927 PSLYPPQDOLRGYQSSKSM-MNLRDVTGGVVRVQAKELMMSQIASQRNSQSGSSHLSS 985

QY 819 VCLDRGLMLK-----PHCTVSFGLSRGGHMS-LMLPGYNQPACKLSHGKT 863
DB 986 T--NTSGRMKMGDMFQOSNNNDWHPSSSTVDI-----HVSVMATSMTPPTQHQYGYR 1037

QY 864 EYGRKLPASEG-----VGKTYGVSRVAVTQHLLSIISLANTL--MSMTNATFFIGHMK 915
DB 1038 DHGTSYPMEDSTHPPPLAPRTTSRVLSSTAPSQAALQIQIQAASSPFRKSDRNSLSRRFLS 1097

QY 916 KGPTRP-----PRPSTPDLSKARG--SPPTSSNI 942
DB 1098 NGGAQPKTVWSPSALANQAPRRSNSNLFPAATNLVEPPTSTNL 1140

RESULT 11
T19550
Hypothetical protein D1054.15 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19550; T20311
R;Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19140
A;Accession: T19550
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-494 <WIL>
A;Cross-references: UNIPROT:Q18295; EMBL:Z73970; PIDN:CAA98247.1; GSPDB:GN00023; CESP:D1054.15
A;Experimental source: clone C29A12
R;Matthews, P.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19255
A;Accession: T20311
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-494 <W12>
A;Cross-references: EMBL:Z74030; PIDN:CAA98448.1; GSPDB:GN00023; CESP:D1054.15
A;Experimental source: clone D1054
C;Genetics:
A;Gene: CESP:D1054.15
A;Map position: 5
A;Introns: 22/1; 61/2; 249/3; 475/3
C;Superfamily: PR11 protein; WD repeat homology

Query Match 2.4%; Score 189; DB 2; Length 494;
Best Local Similarity 24.4%; Pred. No. 4e-05;
Matches 79; Conservative 36; Mismatches 143; Indels 66; Gaps 12;

QY 398 SVIPNSNEPLKVTASVYIPAHGRLVCGREDGS-----IVIVPATQTAIVOLLQGEHMLR 451
DB 106 SVIASAGTFLAIT-----DGSGKLIV-NQOQSAKSGTLLPLPLGNSS-----KGEDNTT 154

QY 452 RG-----WPHRTLGRHNRKVTCLLYPHQVSARYDORYLISGGVDFSV 494
DB 155 RLLSPSKAPMMKPKWHPKLYRVASGHTGV-----RAVDVEPCNWFASGGADRII 208

QY 495 IIVDIPSGEMKHI FCVHGGEITQLLVPENCARSARVQHICSVASDHSVGLLSLRKCKICM 554

Db 209 KIWLASGQLKLSITGHISVRVAVKVSPRH-----PFLFGSGEDQVKWCWLDYKNVIR 262
QY 555 LASRHLPPIQVIKWRSDDDVYVGCSDGVSYYVQMDTGALDRCMGITAVEILNACDEAV 614
Db 263 HYHGLSVAQALSVPFSLDVLVTCARDSTARVMDRTKAQVHCFAGHTNTVADVVCQSDV 322
QY 615 PAAVDSLSHPAVNLKQAMTRSLAALKNMMAHKLQTLA-----TNLLASEASD---KGN 665
Db 323 PQVITASHDATVRLWDLAAGRSWCTL---THHKKSVRALTIHPRLNMFASASPDNIKQWK 379
QY 666 LPK-----YSHNSLMVQAIKTN 682
Db 380 LPKGEFMQNLSGHNAI-INTLSSN 402

RESULT 12
T18526
SRBP cleavage activating protein - Chinese hamster
N:Alternate names: SCAP protein
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18526
R:Hua, X.; Nohturfft, A.; Goldstein, J.L.; Brown, M.S.
Cell 87, 415-426, 1996
A:Title: Sterol resistance in CHO cells traced to point mutation in SRBP cleavage-activating protein
A:Reference number: Z18949; MUID:97053781; PMID:8698195
A:Accession: T18526
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1276 <HUA>
A:Cross-references: UNIPROT:P97260; EMBL:U67060; NID:g1675219; PID:g1675220; PIDN:AAB191
A:Experimental source: ovary
C:Genetics:
A:Gene: SCAP
C:Function:
A:Description: regulates cholesterol metabolism by stimulating cleavage of transcription factor SREBP
C:Keywords: transmembrane protein

Query Match 2.4%; Score 188.5; DB 2; Length 1276;
Best Local Similarity 21.5%; Pred. No. 0.00021;
Matches 136; Conservative 82; Mismatches 196; Indels 219; Gaps 34;

QY 51 ELQINRALLFGHTASTITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLACT 110
Db 764 ETEIVP-LVLRGHLMDIECL-----ASDGMLLVSCCLAGQVCWDAQTGDL--TRIP-- 813
QY 111 HTGIQYQFVSGNQREGRLCHGYPEILLVVDATSLVLYSLVS---KISPDWISSMSII 167
Db 814 -----RPGSRDS---CGG-----GAFQENWERLSDGKTSPEEPGESPL 853
QY 168 RSHRTQEDTVVALSVTGILKWIWTVSEISDMQTEPIFEEESKPIYQNCQISFCFTQ 227
Db 854 R-HRPRGPPQALFGQPDPOLITCLIDTNFVSQLPPEP---TQPEPRHAGC----- 899
QY 228 RSLLVVCSKWRVPDAG-DYSLCSPSENGQWTGDFVSSDKVLIWENGOSYIYKLP 286
Db 900 -----GRADSGYDFSLVR-----VYQEEGLAAV-RMP 928
QY 287 ASCLPASDS-----FRSDVGKAVENLIPVQHILLDRKDKELICPPVTRFFYGCREFYHK 342
Db 929 ALRPPSPGSLPOASQEDGAPEKSGPP-----LAWAPSTAGSIWLSLQGN 975
QY 343 LLIQGSSGRINTWISDTADKQGSSE---GLAMTTSISLQEAFLKNCPCAGIIDQLSV 399
Db 976 LIVVGRSSGRLEWDAIEGVLCSDNDEVSSGITLVFLDRIRIVARLN-----GSLDFSL 1031
QY 400 IPNSN-EPLKV-----TASVY-----IP-----AHGLVC 423
Db 1032 ETHISLSPQFRGTPGSGSPSSSVSSNTVACHLTHVPCAHQKPYALRAAAGRLVT 1091
QY 424 GREDSIVIVPATQTAIVQLLOQHEMLRRGWPPHRTLGRHNKVTCLLYPHQVSARYDQR 483

Db 1092 GSQDHLTRLVRLEDSCL-----FTLQHGSAIT-TVYIDQTMV----- 1129
QY 484 YLISGVDFSVLIWIDFSG-EMKHIFCVHGHEITOLLVPPENCARSVQHICISVAS- 538
Db 1130 -LASGGQGAICLMDVLTGSRVSHTF-AHRGDTVSL-----TCTTS-----CVISSGLDD 1177
QY 539 -----DHSVGLLSLREKKCICMLASRHLFPIQ-----VIKWRPSDDYLVVGCSDGS 583
Db 1178 LINIWRSTGI-----XLYSIQDLGGASLGI-----SDNLLVTG-GQSC 1218
QY 584 VYVQMDTGALDRCMGITAVEILNACDEAVPA 616
Db 1219 VSFMDLNYGDLLOTV-----YLGKNSEAQPA 1244

RESULT 13
T18521
beta transducin-like protein - Podospora anserina
C:Species: Podospora anserina
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18521
R:Saupé, S.; Turcq, B.; Begueret, J.
Gene 162, 135-139, 1995
A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
A:Reference number: Z18944; MUID:96009891; PMID:7557402
A:Accession: T18521
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1356 <SAU>
A:Cross-references: UNIPROT:Q00808; EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775
C:Genetics:
A:Gene: het-el
A:Introns: 761/3

Query Match 2.4%; Score 186; DB 2; Length 1356;
Best Local Similarity 18.4%; Pred. No. 0.00036;
Matches 113; Conservative 76; Mismatches 220; Indels 206; Gaps 22;

QY 22 ISAVLLTDDGATVTCGHDGQICLWLSVELQINPRALLFGHTASTITCLSKACASSDKQY 81
Db 844 VLSVAFSADGQVRVSGSDDKTIKIWDTA---SGTQTTLGHHGGSVMSVA---FSPDRR 897
QY 82 IVSASEGEMCLWDVSDGRCIEFTKLACTHTG-IQFYQFVSGNQREGRLCHGYPEILL 140
Db 898 VASGDDDKTIKIWDAAAGTC---TQTEGHGRVQSVAFSPDQR----- 939
QY 141 VDATSLEVLVSLVSKISPDWISSMSIIRSHRTQEDTVVALSVTGILKWIWTVSEISDMQ 200
Db 940 -----VASGDDDKTIKIWDAAAGTC--- 959
QY 201 TEPIFEEESKPIYQNCQISFCFTQSRLLVVCVKYWRVFDAGDYSLCSPSENGQWTW 260
Db 960 TQTEGHGSSVL-----SVAFSPDQRVSGSDDKTIKIWDTASGTCTQTEGHGGSVM 1013
QY 261 T-----GDFVSS---DKVI-IW-TENGOSYIYKLPASCLPASDSFRSDVGKAVENLI-- 308
Db 1014 SVAFSPDQRVSGSDDKTIKIWDATAG-----TCTQTEGH-----GQWQSVFVS 1060
QY 309 PPVQHILLDRKDKELICPPV---TRFFYGCREFYHKLLIQGSSGRINTWISDTADK 364
Db 1061 PDGQVRVSGSDDKTIKIWDVSGTCTQTEG-----HGDS-----VMSVAFSPDG 1105
QY 365 QGSEGLAMTTSISLQEAFLKNCPCAGIIDQLSVIPNSNEPLKVTASVYIAPHGLVCG 424
Db 1106 Q-----RVASG 1111
QY 425 REDGSIVIVPATQTAIVQLLOQHEMLRRGWPPHRTLGRHNKVTCLLYPHQVSARYDQR 484
Db 1112 SIDGTIKIWDAAAGTCCTQTEG---HCGW-----VHSVAFSPDQR 1149
QY 485 LISGVDFSVLIWIDFSGEMKHIFCVHGHEITOLLVPPENCARSVQHICISVASDVGL 544
Db 1150 VASGISDGTIKIWDAAAGTCCTQTEGHGGSVQSVAFSPDG-----QVRVSGSDDKTIKI 1203

QY 545 LSLREKCIIMASRLHPPIQVIRSDDDYLVVCGSDGVVYVQMDTGALDRCMGITAV 604
DB 1204 WDTASGTCTQTLRGHGQVQVAFSPDQGRVASGSDNTIKIWDTSAG-----TCT 1254
QY 605 EILNACDEAVPAVD 619
DB 1255 QTLNVGSTATCLSPD 1269

RESULT 14
A96638
hypothetical protein Flp17.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
R:Accession: A96638
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96638
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1194 <STO>
A:Cross-references: UNIPROT:O22725; GB:AE005173; NID:g2443881; PIDN:AAB71474.1; GSPDB:GN
A:Gene: Flp17.7
A:Map position: 1

Query Match 2.3%; Score 182.5; DB 2; Length 1184;
Best Local Similarity 17.3%; Pred. No. 0.0005;
Matches 217; Conservative 180; Mismatches 415; Indels 451; Gaps 54;

QY 342 KLLIQDSSGRILNWNISDTADQSGE-----EGLAMTTSISLQAFDK-----LNPCP 390
DB 29 RLFTGGDDYKYNLMAIGKPTSLMKNDATAPYQWGLGHTSAVDSVAFSAEVLVLAGAS 88
QY 391 AGHIDOLSVIPNSELKV-----TASVYIPAHGRVCGREDGSIIVPATQTA 439
DB 89 SGVILKWDV-----EAKNVRAFTGHRNSCAVEFHPFGFEFLASGSSDANLKIWDIRKKG 143
QY 440 IVOLLOGEHMLRGWPHRTLRGRNKVTCLLYPHQVSARYDORYLISGVDPFSVIWDI 499
DB 144 CIQ-----TYKHSRGISTIRFT-----DGRWYVSGLDNVVKVMDL 181
QY 500 FSGEMKHFVCHGGEITQLVPPENCARSVOHCICSVASDHSVGLLSREKKCIMLASRH 559
DB 182 TAGKLHFEKHFEGPIRSLDFF-----LEFLATGSADRTVKFWDLETFELIGSTRPE 235
QY 560 LPPIQVIRKWRPS-----DDYL-----VYVGS- 580
DB 236 ATGVRSIKHPDGRITFLCGLLDSLKYVMEPVVCHVDGMGWTGLDGLTCEGKLACSV 295
QY 581 -DGSVVVWQMD-----TGALDR-CWNGITAVAILNACDE----- 612
DB 296 YQNSVGIVWSDISQIBFYGTGSADKKECV-----EKILSALDQSPDRIKSTPRSSSPD 350
QY 613 -----AVPRAVDSLSHPAVNLKQAMTRSLAALKKNWAHKLQTLATN 654
DB 351 YETKEIKNIYIDSTGNSAVALHSGSLSTPATSTGQGNKSLV-----HSVVRPDS 404
QY 655 LLASEASDKGN-----LPKY-----SHNSLMVQAIKNLTDP--- 686
DB 405 -IGKDSDDSGKESITFSKTPGMLRPAYVRKTPTKFDETKQSVAGVILKSGLDGEKK 463
QY 687 -DIHVLF-----FDVEALIIQLLT---EASRPNT-----ALISPENLQKA--- 723

DB 464 LDTETAFDSEMGSRNPYDADDSIIKSIITNKFQALLPESPTDEAKCMLKPPRVQKSPST 523
QY 724 -----SGSDKGGSPFLTGRAAVLFQVKE-----TIKENIK-----EHL 758
DB 524 KYNEARWATSDSGALDSKKNGLSSRDMDLPTGLADDRGNSNCFEEDENKSISSRSERV 583
QY 759 LDDEEDBEIMQRRESDP-----EYRSSKSKPIUTLE--YNLTMDTAKLPMSCILHAWGLN 813
DB 584 LSPKAGDELKKSRSRSGSNHFIIGYEFVYMQALNPLEDMRHLVIDSLK----- 631
QY 814 EVLDEVCLDRIGMLKPHCTVSGLLSRGGHMSLMLPGYNQPACKLSHGKTEVGRKLPASE 873
DB 632 -----TAVFAKVVSGRTRSLVERFERGE-KITHSEDKA-----ASA 667
QY 874 GVGKGTGYVSRAVTTQHLISLANTLSM-----TNATFFIGDHMKKGTPRPRSTPD 928
DB 668 TVVHSSNSVEERPLTASVQTVSMMPVQVMPVKLDQATNSTTV-----DVPV 713
QY 929 LS--KARGSPPTSSNIVQGIQVAA---PVVSARSDADHSGSDPPSPAPALHTCFLVNEG 983
DB 714 LSTRRTKSTFVRVMPVVLGRDTSMATDTPPVTSITRDRTSATN----- 756
QY 984 WSQLAMHCMLPDLGLDKFR-----PILLEMLARR-----NQDR--CL 1021
DB 757 -----LTSDSVSGVTSKQRTTSPAPVMPMLNQTTMKSDPSITSTWPDRTSAT 806
QY 1022 EVREAAQALLAELRRIEQGRKEAIDAWAPYLPQYIDHVI-----SPGVTS-----EAA 1071
DB 807 DLTSVDSVSGVSSRQRTSP-----APVPMKLNQTKIKSDEPITSTRDRPSA 856
QY 1072 QYITT--APDAGSPEAKV-----QEEHDLVDDD-----ITTCGLSSVPO- 1109
DB 857 TNLTSDESPVTSQAKTSPAPVTPVILNQRTTNKMSDEPPVISTRLRTSSARVMPVI 916
QY 1110 MKKISTSYER-----RKQATAIVLGVGAEGASIEBPBKLTTRSSS----- 1154
DB 917 LNOASTYDERPLSSRSRARTSPARIMPKLNQADNMPSEYEPVALTRSARNSPARVIVP 976
QY 1155 QIPEGFGLTSGGSN-----YSLARHTCKALTF-----LLLOPPSP 1189
DB 977 KLNQATNVTADASHIRSRORFPTQTLATPAVFDQVDMTLDDETTKQSSDILTQKEEP 1036
QY 1190 KLPPHTIRRTAIDLGR--GFTWPEYMDVSAVLMGLLELCAEAKQLANITMGLPLSP 1247
DB 1037 QIS-----GREDDGDIWEILMRTHSEVLNLT-----QSRLTKQLVAVFGT 1077
QY 1248 ADSARSARHALSLIATAPPAFITTIKAEVH-RHTALAANTQSQNMHTTILAR----- 1301
DB 1078 VIQSTVSARVVGVVDLHAEERYLYFRLLYLVHVTIISLFRESKTNRELQVTLPLKFCGI 1137
QY 1302 AKG 1304
DB 1138 AKG 1140

RESULT 15
S59317
DIP2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L3116; protein YLR129W
C:Species: Saccharomyces cerevisiae
C:Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 16-Aug-2004
C:Accession: S59317; S64971; S64966; S69416
R:Delius, H.
submitted to the EMBL Data Library, June 1995
A:Description: 36.8 kb of S.cerevisiae chromosome XII including ACE2, CK11, PDC5, SLS1
A:Reference number: S59317
A:Accession: S59317
A:Molecule type: DNA
A:Residues: 1-943
A:Cross-references: UNIPROT:Q12220; EMBL:X91258; NID:g995686; PIDN:CAA62640.1; PID:g995
A:Experimental source: strain S288C
R:Delius, H.; Hebling, U.

submitted to the Protein Sequence Database, May 1996

A;Reference number: S64967
A;Accession: S64971
A;Molecule type: DNA
A;Residues: 1-943 <DEW>
A;Cross-references: EMBL:Z73301; NID:g1360538; PIDN:CAA97699.1; PID:e245568; PID:g136053
A;Experimental source: strain S288C
R;Verhasselt, P.; Voet, M.; Volckaert, G.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64943
A;Accession: S64966
A;Molecule type: DNA
A;Residues: 1-458 <VER>
A;Cross-references: EMBL:Z73301; MIPS:YLR129w
A;Experimental source: strain S288C
R;Verhasselt, P.; Volckaert, G.
submitted to the EMBL Data Library, September 1995
A;Reference number: S69393
A;Accession: S69416
A;Molecule type: DNA
A;Residues: 1-458 <VEW>
A;Cross-references: EMBL:X89514; NID:g1297019; PIDN:CAA61707.1; PID:e198829; PID:g129704
C;Genetics:
A;Gene: SGD:DIP2
A;Cross-references: SGD:S0004119; MIPS:YLR129w
A;Map position: 12R
C;Superfamily: WD repeat homology
F;117-150/Domain: WD repeat homology <WD1>
F;200-231/Domain: WD repeat homology <WD2>
F;387-419/Domain: WD repeat homology <WD3>
F;569-602/Domain: WD repeat homology <WD4>

Query Match 2.3%; Score 182; DB 2; Length 943;

Best Local Similarity 20.2%; Pred. No. 0.00037;

Matches 139; Conservative 96; Mismatches 239; Indels 214; Gaps 33;

QY	10	IVLWGRKAPT-----HCISAVLLTDD--GATIVTGHCHDQICLWDSVLELQINPRALL	60
Db	103	IKVWDLMSKTVLNFNGHKAATYLLQPDGTGTRLSGSKDSNIIVWDLVGEVGLYK---L	159
QY	61	FGHTASITCLSKACASDKQYIVVSASESGEMCLWDSVSDGRCIEFTKLACTHTG-----	113
Db	160	RSHKDSIT--GFWCQGED--WLSTSKDGMIKLWDLKTHQICB-THIA--HTGECWGLAV	212
QY	114	-----IQPYQSVGNQREGRLCHGHYHEILVVDATS-----LEVLYSLVS	154
Db	213	KDDLITTTGTDQSVKIKWLDIENDKMG-----GKLTEMGIFEXQSKQKGLKIEFITNSD	267
QY	155	KISPDWISSM-STIRSHRTOEDTFWALSVTGILKWIWITSEISDMQDTPPIPEEESKPIY	213
Db	268	KTSTFFIQNADKTIETFRIRKEIEA---RGLKK---REKRLKEKGLTE---BEIAKSI-	317
QY	214	CONCQISFCFAF-----TORSLLVCSKYVRVFDAGDYSLLCSPSENGQTTWGGDFVS	267
Db	318	-----KESYSSFILHPFQTIRSLVKIKSASWTT-----VS	347
QY	268	SDK--VLIWTENGQSYIKL-----PACSLPAS-----DSFRSDVGKAVENLIIPVQH	313
Db	348	SSKLELVLTSSNTIETIYISPIEYKRDPTSPAPLKTHITELQOQTOTDVS-----	396
QY	314	ILLDRDKKELLICPPVTRFPYCGREYFHKLIQGDSSGRINIWNISDTADKQSGSEGLAM	373
Db	397	--IDISDD-----NKLATA-SNGSLKIWNKTHKICIRTFECGYAL	434
QY	374	TTSISLQEAFLKLNPCPAGII-----DQLSVIPNSNEPLKVT-----SVYIPA	417
Db	435	TKCF-----LPGGLVILGTNGELQLFDLASSSLDTIEDAHDAAIWSLDLTS	483
QY	418	HG-RLVCGREDGSI-----VIVPATQTAIVQLLOGEH-----MLRGRWPPH	457
Db	484	DGRLVTGSADKTVKFWDFKVENSLVPGTKNFKLPVLKLGHHDTTLELTDILCVRVSPDD	543
QY	458	R-----TLRGRHNKVTCL-----LYPHQ-----VSARYDQRYLISGGVDFSVIIWDIF	500

Db	544	RYLAISLLDNTVKVPFLDSMKFVLSLYGHKLPVLSIDISFDSKMIITSSADKNIKIWGLD	603
QY	501	SGEMKHIFCVHSGEITQLLVPPENCARSVQHCHICSVASDHSVCLLSLRREKKCIIMLASRHL	560
Db	604	FGDCHKSLPAHQDSIMNVKFLPQS-----HNFFSCSDAVVKYWDGKFECEIQKLYAHQ	657
QY	561	FPQIVIKWRPSSDDYLVVVGSGSDGSVVWQ	588
Db	658	SEVWALAVATDGGFVWSSSHDSIRIWE	685

Search completed: June 30, 2005, 17:24:27
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:13:41 ; Search time 48 Seconds
(without alignments)
2317.231 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773
Sequence: 1 MAGNSLVFLVLMGRKAPTH.....TSNRNVILMAHDGKEHRFMV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	174.5	2.2	587	3	US-08-899-578-2
2	166	2.1	2629	2	US-08-751-189-4
3	166	2.1	2629	2	US-09-060-836-4
4	166	2.1	2629	3	US-09-184-445-4
5	163	2.1	1194	3	US-09-092-508-2
6	163	2.1	1194	3	US-09-435-115-2
7	163	2.1	1194	3	US-09-069-023-26
8	163	2.1	1194	4	US-09-098-310-2
9	163	2.1	1194	4	US-09-538-092-825
10	163	2.1	1194	4	US-09-949-016-6030
11	163	2.1	1196	4	US-09-949-016-10065
12	163	2.1	1196	4	US-09-949-016-10066
13	163	2.1	1205	3	US-09-092-508-16
14	163	2.1	1205	3	US-09-435-115-16
15	150.5	1.9	499	3	US-08-988-856B-4
16	150.5	1.9	499	4	US-09-843-845-4
17	150.5	1.9	499	4	US-09-917-254-60
18	150.5	1.9	499	5	PCT-US95-01806-4
19	150.5	1.9	1237	4	US-09-949-016-6842
20	150.5	1.9	1239	4	US-09-949-016-10063
21	150.5	1.9	1239	4	US-09-949-016-10064
22	150.5	1.9	2627	2	US-08-751-189-3
23	150.5	1.9	2627	2	US-09-060-836-3
24	150.5	1.9	2627	3	US-09-184-445-3
25	149.5	1.9	512	4	US-09-949-016-7770
26	148	1.9	486	4	US-09-722-971-21
27	147.5	1.9	527	4	US-09-949-016-9616

28	147	1.9	430	4	US-09-722-971-4	Sequence 4, Appli
29	146	1.9	542	4	US-09-832-161-16	Sequence 16, Appl
30	145.5	1.9	943	4	US-09-949-016-6758	Sequence 6758, Ap
31	145.5	1.9	955	4	US-09-949-016-10122	Sequence 10122, A
32	144.5	1.9	445	3	US-08-083-945C-2	Sequence 2, Appli
33	144.5	1.9	445	3	US-08-083-945C-7	Sequence 7, Appli
34	144	1.9	478	4	US-09-722-971-6	Sequence 6, Appli
35	144	1.9	690	4	US-09-291-170A-2	Sequence 2, Appli
36	144	1.9	690	4	US-09-724-884-2	Sequence 2, Appli
37	144	1.9	690	4	US-09-724-592-2	Sequence 2, Appli
38	143.5	1.8	253	4	US-09-724-592-11	Sequence 11, Appl
39	143	1.8	448	4	US-09-722-971-10	Sequence 10, Appl
40	143	1.8	514	1	US-08-190-802A-66	Sequence 66, Appl
41	143	1.8	514	3	US-08-477-346-66	Sequence 66, Appl
42	143	1.8	514	3	US-08-477-089-66	Sequence 66, Appl
43	143	1.8	514	4	US-08-487-072A-66	Sequence 66, Appl
44	142.5	1.8	253	4	US-09-291-170A-10	Sequence 10, Appl
45	142.5	1.8	253	4	US-09-724-884-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-899-578-2
; Sequence 2, Application US/08899578
; Patent No. 6087153
; GENERAL INFORMATION:
; APPLICANT: Greenwald, Iva
; APPLICANT: Hubbard, E. Jane
; TITLE OF INVENTION: SEL-10 AND USES THEREOF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,578
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/53200/JPW/AKC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 278-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-899-578-2

Query Match 2.2%; Score 174.5; DB 3; Length 587;

Best Local Similarity 20.0%; Pred. No. 3.9e-07; Indels 91; Gaps 13;
Matches 67; Conservative 52; Mismatches 125;

QY 336 CREYFKLLIQDSSGRNLNWNISDTADKQSGEEGLAMTTSISLQBAF--DKLNPCPAGI 393

Db 261 CMQIHDDVLVTGSDDDNTLKVWCI-----DKGEVMTLVGHTGVWTSQISQCGRYI 311

QY 394 IDQLSVNSNPLKVTASV-----YIPAHGRLV-CGREDSIVLVPAQTQA--IVQLL 444

Db 312 VS-----GSDTRVTKVMSTVDGSLHTLQGTSTVRCMAGSILVTGSRDRTTLRVMDVE 366
Qy 445 QGEHMLRRGWPHRTLEGRNKVTCILYPHQVSRVDORVLISGVDFSVIINDISGEM 504
Db 367 SGRHL-----ATLGHHAARVCV-----QPDGTTVSGGYDFTVKIWAHTGRC 410
Qy 505 KHIFCVHGGBITOLLVPPEN---CSARV-----QHCIC----- 534
Db 411 IRTLTHNNRVYSLPFESERSIVCSGLDTSIRVDFTRPREGQECVALLQCHTSLTSGMQ 470
Qy 535 -----SVASDHSVGLLSREKCKIMLASRHLFPIQVIKWRPSDDYLVVGCSDGVSYYW 587
Db 471 LRGNILVSCNADSHVRVMDIHEGTCVHMLSGHSAITSLQWP-GRNVATSSDDGTVKLW 529
Qy 588 QMDTGALDR-----CWNGITAVEILNAC 610
Db 530 DIERGALIRDLVTDGSGNGGCIWRLCSTSTMLAC 564

RESULT 2

US-08-751-189-4
; Sequence 4, Application US/08751189
; Patent No. 5919656
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,189
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleksi, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-751-189-4

Query Match 2.1%; Score 166; DB 2; Length 2629;
Best Local Similarity 21.2%; Pred. No. 4.7e-05;
Matches 143; Conservative 88; Mismatches 235; Indels 210; Gaps 40;
Qy 12 LMGKAPTHCISAVLLTDDGATVTCGCHQICLWLSVLEQLNPRALLFGHTAS----- 66
Db 2065 LRHGEVPCVCCS---FSPDGGILATAGRDRLNLCWDMKI-----AQAPLLIHTFSSCHRD 2116
Qy 67 -ITCLSKACASSDKQYIVSASBEGMCLWDVSDGRCIEFTKTLACTHTGTGIFYQFS----- 120
Db 2117 WIT-----GCATWKDNLVSCSDGSGVLWNPEAGQ-----QLGQFSGHQSA 2158
Qy 121 -----VGNQREGRLLCHGHVPEILVVDATSLVLYSVLSKIS-----PWISS 163
Db 2159 VSAVVAVEEHIVSVRDGTLKWDHQG-----VELTSIPAHSGPIISQCAAALEPRPGQPG 2214

Qy 164 MSIIRSHRTOEDTVVALSVTGILKVM--IVTSEISDMQD-----TEPIFEESKPIYCON 216
Db 2215 SEL-----VTVGLDGTATKWLHPLLVQCIRTLQHGSGPVTAAASEASGLLTSD 2265
Qy 217 COSI9FC-----AFTORSLLVCSKYRVRFDAGDYSLLCSPSENGQ--TWGQDFV 266
Db 2266 DSSVQLWQIPKEADDSYKPRSSVAITAVAW---APDGSMMVVG-NEAGELTLWQQAQAV 2320
Qy 267 SSDKV-----IIV-----TENGQSYIYKLPASCLPASDS-----FRSDVGKA 303
Db 2321 ATAQAQPGRVSHLIWYSANSFFVLSEANENVSEWQVGLRKGSTSTSSSLHLKRVLQEDWGL 2380
Qy 304 VE-NLIPPVQHILLDRKKEKELICPPVTRFFYGCREY-FHKLLI-----QGDS 349
Db 2381 TGLGLAPDQGSILMKEDVELLEMKPGSIPSSICRRYGVHSSILCTSEYGLFYLOQGD- 2439
Qy 350 SGRLEINWISDTADKQSGEGLAMTTSISLOBAFQKLNPCPAGIIDLQLSVIPNSNEPLKV 409
Db 2440 SGLLSILEQESGE---PEEILDFNLNLN-----NP-----NGSPVSI 2474
Qy 410 TASVYIPAHGRVLCGREDGSIVIVPATQTATVQLQGEHMLRGGW-----PPHRTLRGH 463
Db 2475 T-QAKPESESLLCATSDGMLNWLSECTS-----EGEWIVDNIWQKAKKPKTOTLETE 2527
Qy 464 RNVKVTCLLYPHQVSARYDQRYLISGVDPFSVIWDIFSGEMKHIFC--VHGGEITOLLVP 521
Db 2528 -----LSPH-----SELDFSIDCW-IDPTNLKAQCKKHLHLSVTALHVL 2566
Qy 522 PENCARSARVQHCICSVASDHSVGLLSIREKKKIMLASRHLF-----PIQVIK-W-RPSDD-Y 574
Db 2567 P-----GLLVTASKORDV---KLWRPSSQLLG--LFRCEGPPVSCLEPNWEPSPPLQ 2613
Qy 575 LVVGCSDGSVVV--WQ 588
Db 2614 LAVGDTQGNLYFLSWE 2629

RESULT 3

US-09-060-836-4
; Sequence 4, Application US/09060836
; Patent No. 5981707
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,836
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,189
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleksi, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

LENGTH: 2629 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-060-836-4

Query Match 2.1%; Score 166; DB 2; Length 2629;

Best Local Similarity 21.2%; Pred. No. 4.7e-05;
Matches 143; Conservative 88; Mismatches 235; Indels 210; Gaps 40;

QY 12 LWRKAPTHCISAVLLTDDGATVTCGHDGQICLDLWLSVELQINPRALLFGHTAS-----66
DB 2065 LRHGEFVCCS---FSPDGGILATAGDRNLLCWDNMI-----AQAPLLIHTFSSCHRD 2116

QY 67 -ITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLAETHGTGIFYOFS-----120
DB 2117 WIT---GCAWTKDNILVSCSDSGVGLWNPAGQ-----QLGQFSGHQA 2158

QY 121 -----VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKIS-----PDWISS 163
DB 2159 VSAVVAVEEHIVSVSRDGLKVNDHQG---VELTSIPAHSGPISQCAAALEPRPGQPG 2214

QY 164 MSIIRSHRTQEDTVALSVTGILKW---IVTSEISDMQD-----TEPIFEESKPIYCON 216
DB 2215 SELL-----VVTVGLDGATKLMHPLLVQCIQTLOGHSGPVTAAASEASGLLTSD 2265

QY 217 CQGISFC-----AFTQSLVAVCSKYWRVFDAGDYSLLCSPGSENGO--TWTCGDFV 266
DB 2266 DSSVOLWQIPKEADDSYKPRSSVAITAVAM---APDGSVMVSG-NEAGELTLWQQA 2320

QY 267 SSDKV-----IIW-----TENQSYIYKLPASCLPASDS-----FRSDVGKA 303
DB 2321 ATAQAPGRVSHLIWYSANSPFVLSANENVSEWQVGLRKSGTSTSSSLHLKRVLQEDW 2380

QY 304 VE-NLIPVQHILLDRKDKELLICPPVTRFFYGCYV-FHKLII-----QGDS 349
DB 2381 TGLGLAPDQGSLLMKEDVLEMKPGSIPSSICRRYGVHSSILCTSKYGLFYLOQGD- 2439

QY 350 SGRINLWNIISDTADKQSGSEGLAMTTSISLQEAFLKPNCPAGIIDQLSVIPNSPLKV 409
DB 2440 SGLLSILEQESGE---FEIILDFNLN-----NP-----NGSPVSI 2474

QY 410 TASVYIPAHGRVCGREDGSIIVPATQTAIVQLQGEHMLRGW-----PPHRTLGRH 463
DB 2475 T-QAKPESSESLLCATSDGMLNLSCTS-----EGEIVDNIWQKAKKPKTQLETE 2527

QY 464 RNKVTCLLYPHQVSARDYRLISGGVDPSVLIWDIFSGEMKHIFC--VHGGEITQLLVP 521
DB 2528 -----LSPH-----SELDFSIDCW-IDPTNLKQAQCKKHLSVTLHVL 2566

QY 522 PENCARSVQHCISVADSHVGLLSLREKKICMLASRLH---PIQVIK-W-RPSDD-Y 574
DB 2567 P-----GLLVTASKDRV---KLWERPSMQLLG--LPRCGPVSCLPEWEPSPLOQ 2613

QY 575 LVVGCSDGSVYV--WQ 588
DB 2614 LAVGDTQGNLYFLSWE 2629

RESULT 4

US-09-184-445-4

Sequence 4, Application US/09184445

Patent No. 6174703

GENERAL INFORMATION:

APPLICANT: Harrington, Lea A.

APPLICANT: Robinson, Murray O.

TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein

TITLE OF INVENTION: 1

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Angen, Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/184,445

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/751,189

FILING DATE: 15-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Olecki, Nancy A.

REGISTRATION NUMBER: 34,688

REFERENCE/DOCKET NUMBER: A-433

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2629 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-09-184-445-4

Query Match 2.1%; Score 166; DB 3; Length 2629;

Best Local Similarity 21.2%; Pred. No. 4.7e-05;
Matches 143; Conservative 88; Mismatches 235; Indels 210; Gaps 40;

QY 12 LWRKAPTHCISAVLLTDDGATVTCGHDGQICLDLWLSVELQINPRALLFGHTAS-----66
DB 2065 LRHGEFVCCS---FSPDGGILATAGDRNLLCWDNMI-----AQAPLLIHTFSSCHRD 2116

QY 67 -ITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLAETHGTGIFYOFS-----120
DB 2117 WIT---GCAWTKDNILVSCSDSGVGLWNPAGQ-----QLGQFSGHQA 2158

QY 121 -----VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKIS-----PDWISS 163
DB 2159 VSAVVAVEEHIVSVSRDGLKVNDHQG---VELTSIPAHSGPISQCAAALEPRPGQPG 2214

QY 164 MSIIRSHRTQEDTVALSVTGILKW---IVTSEISDMQD-----TEPIFEESKPIYCON 216
DB 2215 SELL-----VVTVGLDGATKLMHPLLVQCIQTLOGHSGPVTAAASEASGLLTSD 2265

QY 217 CQGISFC-----AFTQSLVAVCSKYWRVFDAGDYSLLCSPGSENGO--TWTCGDFV 266
DB 2266 DSSVOLWQIPKEADDSYKPRSSVAITAVAM---APDGSVMVSG-NEAGELTLWQQA 2320

QY 267 SSDKV-----IIW-----TENQSYIYKLPASCLPASDS-----FRSDVGKA 303
DB 2321 ATAQAPGRVSHLIWYSANSPFVLSANENVSEWQVGLRKSGTSTSSSLHLKRVLQEDW 2380

QY 304 VE-NLIPVQHILLDRKDKELLICPPVTRFFYGCYV-FHKLII-----QGDS 349
DB 2381 TGLGLAPDQGSLLMKEDVLEMKPGSIPSSICRRYGVHSSILCTSKYGLFYLOQGD- 2439

QY 350 SGRINLWNIISDTADKQSGSEGLAMTTSISLQEAFLKPNCPAGIIDQLSVIPNSPLKV 409
DB 2440 SGLLSILEQESGE---FEIILDFNLN-----NP-----NGSPVSI 2474

QY 410 TASVYIPAHGRVCGREDGSIIVPATQTAIVQLQGEHMLRGW-----PPHRTLGRH 463
DB 2475 T-QAKPESSESLLCATSDGMLNLSCTS-----EGEIVDNIWQKAKKPKTQLETE 2527

QY 464 RNKVTCLLYPHQVSARDYRLISGGVDPSVLIWDIFSGEMKHIFC--VHGGEITQLLVP 521
DB 2528 -----LSPH-----SELDFSIDCW-IDPTNLKQAQCKKHLSVTLHVL 2566

QY 522 PENCARSVORHCISVASHDVGSLLSREKKCIMLASHLF-----PIOVIK-W-RPDS-DY 574
Db 2567 P-----GLVTASKDRV---KLWERSMQLLG--LPRCGPVSCLEPWNPEPSPLQ 2613
QY 575 LVVGCSDGSGVYV--WQ 588
Db 2614 LAVGDTQGNLYFLSWE 2629

RESULT 5

US-09-092-508-2
; Sequence 2, Application US/09092508
; Patent No. 6291643
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6291643west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,508
; FILING DATE: 05-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,807
; FILING DATE: 05-JUN-1997
; APPLICATION NUMBER: 60/055,258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal

US-09-092-508-2

Query Match 2.1%; Score 163; DB 3; Length 1194;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;
QY 115 QYQFQSVGNOREGRLCHGYPPELVVDATSLVLSVSKISPDWISSMSIIRSHRTOE 174
Db 328 EYLLKQLNQKQFRIRKSSDYDEALDEANSISV--EMLREDIKDYTDLSIL-----QK 380
QY 175 DTVVALLSVTGILKWIVTISBDMQDTEPIFEESKPIYQNCQCSISFCAFTQRSLLVVC 234
Db 381 DVKVPKVLKIL--WDMETE-----EVEDILQE-----FVNKSL-FC 415
QY 235 SKYWRVFDAGDYSLLCSGSPENGQWTGGDFVSSDKVLIWTENGQSYI----- 282
Db 416 DR-----NGKSPR--YYLHDLQVDFLTERKNCQLOQLHKHKKITQFQ 454
QY 283 -YKLPASCLPASDS-----FRSDVGRAVENLIPPVQHI-- 314

Db 455 RYHQPHTLSPDQEDCMYWNFLAYHMASAKMHKELCALMFSLDWIKAKTELGVPAHLIHE 514
QY 315 -----LLDRKKELLICPPVTRFPFYGCREYFHKLIIQGDSSGRLNIWI-----SDT 361
Db 515 FVEYRHILDEKD-----CAVSENF-----QEF-----LSLNGHLGRQFPFNIQGLCEPET 562
QY 362 AD-----KQGEGLAMTTTSLQEAFOKLNPCPAGIIDQLSVINSNEPLKVTAS 412
Db 563 SEVYQAKLOAKQEVNDGMLYLEWINKKNITN-----LSRLVVRH-----TDA 606
QY 413 VYIPAHGRILVCGREBGS-IVIVPATQT-AIVQLLOGEHMLRRGWPPHRTLGRHNRKVTCL 470
Db 607 VY--H--ACFSEDQRIASCADAKTLQVFAETGEKLE-----IKAHEDEVLC 652
QY 471 LYPHQVSARYDORYLISGVDVSFVIWIDIFSEGMKHFVCHGGEITQLLVPPENC---SA 527
Db 653 AF-----STDDRFIATCSVDKVKIWNMSMTGELVHTYDEHSEV-----NCCHFTN 698
QY 528 RVQHCICSV-ASDHSVGLLSREKKCIMLASRHLPIQVIKWRPSDDYLWVGCs-DGSVY 585
Db 699 SSHLLLATGSSDCFLKMDLNQKCRNTMFGHTSVNHCRRFSP-DDKLLASCADGTLK 757
QY 586 VMQMDTGALDRCVMGITAVEILNACD--EAVPAADVLSHPAVNLKQAMTRSLAALKM 643
Db 758 LWDA-TSANERKSINVKQF-FLNLEDPQEDMEVIVKCCSWSADG-----ARIMVAAKN- 808
QY 644 AHKLOTLATNLLASEADKGNLPKYSHNSLWQAKTNLTDPD----- 687
Db 809 ---KIPLMNTDSRSKVADCRGHL-SWVHGVMFSPDGSSFLTSSDDQTIRLWETKVKCKNS 864
QY 688 -----IHLVDFDVEALI-----IQLLTEASRPNTALISPENLQK 722
Db 865 AVMLKQEVDDVFOENEVMVLAVDHRRLQINGRQIDILTE--AQVSCCLSPHLQVI 922
QY 723 ASGSSDKGGSF---ITGKRAAVLFO---QVKET---IKENKEHLDDDEEDEE--- 767
Db 923 AFG--DENGAIIEILELVNRR---IFQSRFOHKKTVMHIQFTADEKTLISSDDAEIQVWN 977
QY 768 -----IMQRRESDPEYRSKSKPLTLEVLNLTMDTAKLPMSCLHANGL---NEVLDE 818
Db 978 WQLDKCIFLRGHQETVKDFRLKNS--RLLSWS-----FDGTVKVWNIIITGNKEKDF 1027
QY 819 VCLDRGLMKPHCTVSFGLLSRGHMSLMLPGVNOPACKLSHGKTEVGRK----- 868
Db 1028 VC-----HQTIVL-----SCDISHATKFSSTADTKAKIWS 1059
QY 869 ----LPASEGVKGTYGVSRAVTTQHLLSIIISLANTLMSMTNATFGIDHMKKGPTPRRP 924
Db 1060 FDLLLPLHE--LRGHNGCVRC-----SAFVSVDSTLLA-----TGD----- 1092
QY 925 STFDLSKARGSPPTSSNIVQGOIKQVAAAPVVSARSADADHSG 965
Db 1093 ---DNGEIR-----IWNVSNGLLHLHLCAP-LSEGAATHGG 1124

RESULT 6

US-09-435-115-2
; Sequence 2, Application US/09435115
; Patent No. 6346607
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6346607west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/435.115
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/092,508
FILING DATE:
APPLICATION NUMBER: 60/055,258
FILING DATE: 07-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kettelberger, Ph.D., Denise M
REGISTRATION NUMBER: 33,924
REFERENCE/DOCKET NUMBER: 11669.6USU1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1194 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-435-115-2

Query Match 2.1%; Score 163; DB 3; Length 1194;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;
QY 115 QYQFVGVNREGELLCGHYPETLVVDATSLVLSYLSKISPDWISSMSIIRSHRTOE 174
DB 328 EYLLKQNKQFKIRKSSDYDEALDEAMSISV--EMLRDIDKYDYLDSIL-----QK 380
QY 175 DTVALSVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQISFCFAFTORSLLVVC 234
DB 381 DVKVPVKVLCIL--WDMETE-----EVEDILOE-----FVNKSL-FC 415
QY 235 SKYWRVFDAGDYLSCGSPENGQWTGDFVSSDKVIITWENGQSVI----- 282
DB 416 DR-----NGKSPR--YYLHDLQVDFLTCKNCSQLODLHKIITQFQ 454
QY 283 -YKLPASCLPASDS-----FRSDVGKAVENLIPVQHI-- 314
DB 455 RYHQPHTLSPDQDCMWTYNFLAYHMASAKMKELCALMFSLDWIKAKTELGVPAHLIHE 514
QY 315 -----LLDRKDKELLCPPVTRPFYGCYREYFHKLLIQGDSGRLNIWNI-----SDT 361
DB 515 FVEYRHILDEK-----CAVSENF-----QBF--LSLNGHLGQPPNIVQLGCEPET 562
QY 362 AD-----KQSGEGLAWTTSISIQEAFDKLNPAGIIDQLSVIPNSBPLKVTA 412
DB 563 SEVVQAKLQAKQEVNGLYLEWINKNITN-----LSRLVVRPH-----TDA 606
QY 413 VYIPAHGRLVCGREDGS-IVIVPATQT-AIVOLLQGEHMLRRGPPHRTLRGHNKVTC 470
DB 607 VY---H---ACFSEDDGRIASCGADKTLQVFKATPGEKLE-----IKAHEDEVLC 652
QY 471 LYPHOVSARYDQRLISGGVDFSVIWIIDIFSGEMKHIFCVHGGEITQLLVPPENC---SA 527
DB 653 AF-----STDDRFIATCSVDKVKIWNMTGELVHTYDEHSEGV-----NCHFTN 698
QY 528 RVQHICISV-ASHSVGLLSIREKKCITMLASRLFPQIVIKWRPSDDYLIVGCS-DGSVY 585
DB 699 SSHLLLATGSSDCFLKLDLNOKECRNTWFGHTNSVNHCRFSP-DDKLLASCADGTLK 757
QY 586 VWQMDTALDRCVNGITAVELIACD--EAVPAADVSLSHPAVNLQAMTRSLAAKNM 643
DB 758 LWDA-TSANERKSNVQF-FLNLEDQEDMEVIVKCCSWSDG-----ARIMVAKN- 808

QY 644 AHKKLOTLATNLLASEASDKGNLPKYSHNSLAWQAIKTNLTDPD----- 687
DB 809 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMPSPDGSSFLTSSDDOTIRLWETKVCCKNS 864
QY 688 -----IHVLPFDVEALI-----IQLTEASRPNTALISPENLQK 722
DB 865 AVMLKQEVDFVFOENEVMVLAVDHIRRLQINGRTGQIDYLTE--AQVSCCCLSPHLYI 922
QY 723 ASGSSDKGGSF---LTGKRAAVLFQ---QVKET---IKENIKEHLIDDEEEDBE----- 767
DB 923 AFG--DENGAIETLELVNR---IFQSRFOHKKTVMHIOFTADEKTLISSDDAEIQVWN 977
QY 768 -----IMQRRESDPEYRSSKPLTJLLEYNLTWDTAKLFMSCLHAWGL---NEVLDE 818
DB 978 WQLDKICIFLRGHQETVKDFRLKNS--RLLSWS-----FDGTVMVMNIITGNKEKDF 1027
QY 819 VCLDRGLMKPHCTVSVFGLLSRGHMSLMLPGYNQPACKLSHGKTEVGRK----- 868
DB 1028 VC-----HOGTVL-----SCDISHDATKFSSTSAKTAKIWS 1059
QY 869 ---LPASEGVGKGTGVSVRAVTTQHLLSIISLANTILMSMTNATFIGDHMKKGPTRP RP 924
DB 1060 FDLILLPLHE--LRGNGCVRC-----SAFSDVSTLLA-----TGD----- 1092
QY 925 STPDLSKARGSPPTSSNIYVQGIQKQVAAVPSARSADHSG 965
DB 1093 ---DNGBEIR-----IWNVSNGELLHLCAP-LSEGAATHGG 1124

RESULT 7
US-09-069-023-26
Sequence 26, Application US/09069023A
Patent No. 6348573
GENERAL INFORMATION:
APPLICANT: Nunez, Gabriel
APPLICANT: Inohara, Naohiro
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS
FILE REFERENCE: UM-0333
CURRENT APPLICATION NUMBER: US/09/069,023A
CURRENT FILING DATE: 1998-04-27
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 26
LENGTH: 1194
TYPE: PRT
ORGANISM: Homo sapiens
US-09-069-023-26

Query Match 2.1%; Score 163; DB 3; Length 1194;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;
QY 115 QYQFVGVNREGELLCGHYPETLVVDATSLVLSYLSKISPDWISSMSIIRSHRTOE 174
DB 328 EYLLKQNKQFKIRKSSDYDEALDEAMSISV--EMLRDIDKYDYLDSIL-----QK 380
QY 175 DTVALSVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQISFCFAFTORSLLVVC 234
DB 381 DVKVPVKVLCIL--WDMETE-----EVEDILOE-----FVNKSL-FC 415
QY 235 SKYWRVFDAGDYLSCGSPENGQWTGDFVSSDKVIITWENGQSVI----- 282
DB 416 DR-----NGKSPR--YYLHDLQVDFLTCKNCSQLODLHKIITQFQ 454
QY 283 -YKLPASCLPASDS-----FRSDVGKAVENLIPVQHI-- 314
DB 455 RYHQPHTLSPDQDCMWTYNFLAYHMASAKMKELCALMFSLDWIKAKTELGVPAHLIHE 514
QY 315 -----LLDRKDKELLCPPVTRPFYGCYREYFHKLLIQGDSGRLNIWNI-----SDT 361

Db 515 FVEYRHILDEK-----CAVSENF-----QEF-----LSLNGHLLGRQPPNIVOLGLCEPET 562
Qy 362 AD-----KQSEBGLAMTTSISLOEAFDKLPCPAGIIDQLSVIPNSNEPLKVITAS 412
Db 563 SEVYQOAKLOAQOEVDNGMLYLEWINKNITN-----LSRLVVRPH-----TDA 606
Qy 413 VYIPAHGRVCGREDGS-IVIVPATOT-AIVOLLOQEHMLRGPWPHRTLGRHKNVTCL 470
Db 607 VI-----H-----ACFSEDGQRIASCADKTLQVFAETGEKLE-----IKAHEDEVLC 652
Qy 471 LYPHQVSARYDORYLISGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENC---SA 527
Db 653 AF-----STDDRFIATCSVDKVKIWSMTGELVHTYDEHSEV-----NCCHFTN 698
Qy 528 RVQHICISV-ASHSVGLLSLRBKCMILASRLHPFPIQVWKRPSPDYLWVCGS-DGSVY 585
Db 699 SSHLLLATGSSDCFLLKWLQKCECRNTMFGHTNSVNHCRFSP-DDKLLASCADGTLK 757
Qy 586 VQWMDTGALDRVCVMGITAVEILNACD--EAVPAADVSLSHPAVNLQAMTRRSIALAKNM 643
Db 758 LWDA--TSANERKSINVQF-FLNLEDPQEDMEVIVKCCSWADG-----ARIMVAAKN- 808
Qy 644 AHKLOTLATNLLASEADKGNLPKYSHNSLMVQAIKTNLTDPD----- 687
Db 809 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMFSPDGSFLTSSDDQTIRLWETKVKCKNS 864
Qy 688 -----IHVLPFDVEALI-----IQLLTFEASRPNTALISPENLQK 722
Db 865 AVMLKQEVDFVFOENEVWVLAVDHRLQLINGRTGQIDYLTE--AQVSCCLSPHLOYI 922
Qy 723 ASGSSDKGGSF-----LTGKRAAVLFO---OVKET---IKENIKHELLDDDEE--- 767
Db 923 AFG--DENGAIETLELVNR---IFQSRFOHKKTWHIQTFADEKTLISSDDAEIQVWN 977
Qy 768 -----IMRQRRESDPEYRSKSKPLTLEYNLTMDTAKLFMSCLHAWGL---NEVLDE 818
Db 978 WQLDKCIFLRGHQETVKDFLLKNS--RLLSWS-----FDGTVKVWNIITGNKEKDF 1027
Qy 819 VCLDRGLMKPHCTVSPGLLSRGHMSLMLPGYNQAPACKLSHGKTEVGRK----- 868
Db 1028 VC-----HOGTVL-----SCDISHDATKFSSTSDAKTIKWS 1059
Qy 869 ---LPASGVGKGTGVSVRAVTTQHLLSIISLANTLMSMTNATFTGDHMKKGTPTPRP 924
Db 1060 FDLLEPLHE--LRHNGCVRC-----SAFVDSSTLLA-----TGD----- 1092
Qy 925 STPDLSKARGSPPTSSNIVQGIQKVAAPVVSARSADHSG 965
Db 1093 ---DNGEIR-----IWNVSGELLHLCAP-LSEGAATHGG 1124

RESULT 8
US-09-098-310-2
; Sequence 2, Application US/09098310
; Patent No. 6403765
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: TRUNCATED APAF-1 AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 480140.438
; CURRENT APPLICATION NUMBER: US/09/098.310
; CURRENT FILING DATE: 1998-06-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1194
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-098-310-2

Query Match 2.1%; Score 163; DB 4; Length 1194;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;

Qy 115 QFQVQSVGNQREGRLLCHGHYPEILLVVDATSLVLYSLVSKISPDWISSMISIRSHRTOR 174
Db 328 EYILQOLQNKQFRIRKSSSYDEALDEAMSISV--EMLREDIKDYTTLSIL-----QK 380
Qy 175 DTVALSVTGILKMWITVSEISDMQDTEPIPEESKPIYQCNQCSISFCAFORSLLVVC 234
Db 381 DVKVPKVLICL--WDMETE-----EVEDILQE-----FVNKSL-FC 415
Qy 235 SKYWRVFDAGDYLSCSGPSENGQWTGQDGVSSDKVIIWTENGQSYI----- 282
Db 416 DR-----NGKSFR--YILHDLQVDFLETKNCSQLQDLHKHKKIITQFQ 454
Qy 283 -YKLPAASCLPASDS-----FRSDVGKAVENLIPPVQHI-- 314
Db 455 RYHQPHLLSPDQEDCMYWNFLAYEMASAKMKELCALMFLSLDWIKAKTELGVPAHLIHE 514
Qy 315 -----LLDRDKKELLICPPVTRFFYGCYREYFKLLIQDSSGRLLMNI-----SDT 361
Db 515 FVEYRHILDEK-----CAVSENF-----QEF-----LSLNGHLLGRQPPNIVOLGLCEPET 562
Qy 362 AD-----KQSEBGLAMTTSISLOEAFDKLPCPAGIIDQLSVIPNSNEPLKVITAS 412
Db 563 SEVYQOAKLOAQOEVDNGMLYLEWINKNITN-----LSRLVVRPH-----TDA 606
Qy 413 VYIPAHGRVCGREDGS-IVIVPATOT-AIVOLLOQEHMLRGPWPHRTLGRHKNVTCL 470
Db 607 VI-----H-----ACFSEDGQRIASCADKTLQVFAETGEKLE-----IKAHEDEVLC 652
Qy 471 LYPHQVSARYDORYLISGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENC---SA 527
Db 653 AF-----STDDRFIATCSVDKVKIWSMTGELVHTYDEHSEV-----NCCHFTN 698
Qy 528 RVQHICISV-ASHSVGLLSLRBKCMILASRLHPFPIQVWKRPSPDYLWVCGS-DGSVY 585
Db 699 SSHLLLATGSSDCFLLKWLQKCECRNTMFGHTNSVNHCRFSP-DDKLLASCADGTLK 757
Qy 586 VQWMDTGALDRVCVMGITAVEILNACD--EAVPAADVSLSHPAVNLQAMTRRSIALAKNM 643
Db 758 LWDA--TSANERKSINVQF-FLNLEDPQEDMEVIVKCCSWADG-----ARIMVAAKN- 808
Qy 644 AHKLOTLATNLLASEADKGNLPKYSHNSLMVQAIKTNLTDPD----- 687
Db 809 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMFSPDGSFLTSSDDQTIRLWETKVKCKNS 864
Qy 688 -----IHVLPFDVEALI-----IQLLTFEASRPNTALISPENLQK 722
Db 865 AVMLKQEVDFVFOENEVWVLAVDHRLQLINGRTGQIDYLTE--AQVSCCLSPHLOYI 922
Qy 723 ASGSSDKGGSF-----LTGKRAAVLFO---OVKET---IKENIKHELLDDDEE--- 767
Db 923 AFG--DENGAIETLELVNR---IFQSRFOHKKTWHIQTFADEKTLISSDDAEIQVWN 977
Qy 768 -----IMRQRRESDPEYRSKSKPLTLEYNLTMDTAKLFMSCLHAWGL---NEVLDE 818
Db 978 WQLDKCIFLRGHQETVKDFLLKNS--RLLSWS-----FDGTVKVWNIITGNKEKDF 1027
Qy 819 VCLDRGLMKPHCTVSPGLLSRGHMSLMLPGYNQAPACKLSHGKTEVGRK----- 868
Db 1028 VC-----HOGTVL-----SCDISHDATKFSSTSDAKTIKWS 1059
Qy 869 ---LPASGVGKGTGVSVRAVTTQHLLSIISLANTLMSMTNATFTGDHMKKGTPTPRP 924
Db 1060 FDLLEPLHE--LRHNGCVRC-----SAFVDSSTLLA-----TGD----- 1092
Qy 925 STPDLSKARGSPPTSSNIVQGIQKVAAPVVSARSADHSG 965
Db 1093 ---DNGEIR-----IWNVSGELLHLCAP-LSEGAATHGG 1124

RESULT 9
US-09-538-092-825
; Sequence 825, Application US/09538092

Qy	362	AD-----KQSEEGAMTTSISLOAEFKLNCPCAGIIDQLSVIPNSNEPLKVTA5	12
Db	563	SEVYQAKLQAKQVEDNGMLYLEINKNIITN-----LSRLVVRPH-----TDA	606
Qy	413	VYIPAHGRLVCGREDGS--IVIVPATQT--AIVOLLOGEHMLRGWPPHRTLGRHNKVCL	470
Db	607	VI-----H-----ACFSEBQRIASCGADKTLQVKAETGEKLE-----IKAHEBVLCC	652
Qy	471	LYPHQVSARYQRYLISGGVPFSVIIWDIFGSEMKHIFCVHGGEITQLLVPPENC-----SA	527
Db	653	AF-----STDDRFIATCSVDKKVKKIWNSMTGELVHTYDEHSEQV-----NCFHFN	698
Qy	528	RVQHCICSV--ASDHSVGLLSREKKCICMLASHPLFPIQVIKWRPDDVYLVCSS--DGSVY	585
Db	699	SSHLLLLATGSDCFKLKWDLNQKSCRNTMFGHTVSNVHCRFSP--DDKLILACSDAGTUK	757
Qy	586	VWQMDTGALDRCVMGITAVEILINACD--EAVPAADVLSLSPAVNLKQAMTRSLAALKNM	643
Db	758	LWDA--TSANERKSINVKQF--FLNLEDQPQEDMEVIVKCCSWSADG-----ARIMVAAKN-	808
Qy	644	AHHKLOTLATNLLASEASDKGNLPKYSHNSLMVOAIKNLTDPD-----	687
Db	809	---KIFLWNTDSRSKVAIDCRGHL--SNWHGVWFPSPDGSSFLTSSDDQTLRLWETKVCVKN5	864
Qy	688	-----THVLFFDVEALI-----IQLTEASRPNTALISPENLOK	722
Db	865	AVMLKQEVVDVVFQENEVMVLAVDHTRRLQLINGRTGQIDYLTE--AQVSCCCLSPHLQVI	922
Qy	723	ASGSSDKGGSF-----LTGKRAVLQF--QVKET---IKENIKHLLDDEEER-----	767
Db	923	AFG--DENGAIIEILVNNR---IFQSRFOHKTKVWHIQFTADERKTLISSDDASIQVNW	977
Qy	768	-----IMQRREESDPEYRSKSPLTILLYNLTMDTAKLFMSCLHAWGL---NEVLDE	818
Db	978	WLODKCIFLRHQETVKDFRLKNS--RLLSWS-----FDGTVKWNIIITGNKEKDF	1027
Qy	819	VCLDRGLMKPCHTYSVFGLLSRGGHMSMLPGYNQPAACKLSHGKTEVGRK-----	868
Db	1028	VC-----HOGTVL-----SCDISHDATKFSSTSDAKTAIWS	1059
Qy	869	---LPASEGVCKGTGYVSRAVTTQHLLSIIISLANTLMSMTNATFIGDHMKKGTPRPPR	924
Db	1060	FDLLLPLEH--LRHNGCVRC-----SAFSVDSTLLA-----TGD-----	1092
Qy	925	STPDL5KARGSPPTSSNIYVQGIKQVAPVVSARSADHSG	965
Db	1093	--DMGEIR-----TWNYSNGELLHLCAP--LSEEGAATHG	1124

RESULT 11

```

US-09-949-016-10065
; Sequence 10065, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10065
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10065

```

Query Match	2.1%	Score 163	DB 4	Length 1196
Best Local Similarity	19.8%	Pred. No. 2e-05		
Matches	198	Conservative 135	Mismatches 314	Indels 354
Gaps	56			
QY	115	QYQFQSVGNQREGRLCHGHPYELVVDATSLVLSYKISPDWISSMIIIRSHRTOE	174	
DB	330	EYTLKQJQKQFRIKRSYDYEAIDEMSISV--EMLRDIDKYDSDL--	382	
QY	175	DTVVALSVTGILKVIWVTGIBISDMQDTEPIFEESKPIYQCNQCSIFCAFTORSLLVVC	234	
DB	383	DVKVPTKVLCIL--WDMETE--EVEDILOE--	417	
QY	235	SKYWRVPDAGDYSLLCSGSENGCQWTCGDFVSSDKVIIWTENGQSVI--	282	
DB	418	DR-----NGKSPR--YYLHDLQVDFLTEKNCQLQDLHKKIITQFQ	456	
QY	283	-YKLPASCLPASDS-----FRSDVGRAVENLIPVQHI--	314	
DB	457	RYHQPHLSPQEDCMVYNYFLAYHMASAKMHKELCALMFLDWIKAKTELVGPAHLIHE	516	
QY	315	-----LLDRKOKELLCPPVTRFFYGCRRYFKHLIIQDSSGRINIWI-----SDT	361	
DB	517	FVEYRHILDEK-----CAVSENF-----QEF--LSLNGHLLGRQFPPIVQJGLCEPET	564	
QY	362	AD-----KQSGEGLATWTSISLOEAFDKLPCPAGIIDLQSVIPNSNEPLKVITAS	412	
DB	565	SEVYQAKLOAKQEVNMGMLYLEWINKNITN-----LSRLVVRPH-----TDA	608	
QY	413	VYIPAHGRLVCGREDGS-IVIVPATOT-AIVQLLOGEHLMLRRGWPHRTLGRHNRKVTC	470	
DB	609	VY--H--ACFSEDDGRIASCCGADKTQVFKAEATGEKLLIE-----IKAHEDEVLC	654	
QY	471	LYPHQVSARVDORYLLSGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENC--SA	527	
DB	655	AF-----STDDRPIATCSVDKKVKIWNMGTSGLVHTYDEHSQV-----NCHFTN	700	
QY	528	RVQHCICSV-ASDHSVGLLSREKKCITMLASRHLFPQVIKWRPSDDYLVVCGS-DGSVY	585	
DB	701	SSHLLLATGSSDCFLKLDNLQKCRNTMFGHTNSVNHCRFSP-DOKLLASCASDGT	759	
QY	586	VWQMDTGALDRCVNGITAVEILNACD--EAVPAVDLSHPAVNLQAMTRSRSLAALKNM	643	
DB	760	LWDA-TSANERKSINVKQF-FLNLEDPQEDMEVIVKCCSWSDG-----ARIMVA	810	
QY	644	AHKLQTLATNLILASASDKGNLPKYSHNSLWQAIKTNLTDPD-----	687	
DB	811	---XIFLWNTDSKVDACGHL-SWVHGVMFPDGSFLTSSDDQTIRLWETKVKCKNS	866	
QY	688	-----IHVLFPDVEALI-----IQLLTSEARPNALLISPENLQK	722	
DB	867	AVMLKQEVQVVFQENWVLAVDHIRRLQINGRTQIDYLTE-AQVSCCLSPHLQYI	924	
QY	723	ASGSSDKGGSF---LTGKBAVLQ---QVKET---IKENIKHELLDDDEE	767	
DB	925	AFG--DENGAIETILELVNRR---IFQSRFOHKKTVMHIIQTADKTLISSDDAE	979	
QY	768	-----IMRQREESDEPYRSSKSLTLLLEYNLTMDTAKLFMSCLHAWGL-----NEVL	818	
DB	980	WQLDKCIFLRGHQETVKDFELLKNS--RLLSWS-----FDGTVKVMNIITGNKE	1029	
QY	819	VCLDRGLMKLPKCTVSPGLLSRGHMSMLPGYNQPACKLSHGKTEVGRK-----	868	
DB	1030	VC-----HOGTVL-----SCDISHDATKFSSTSDAKTIWS	1061	
QY	869	----LPASGEGVGTGVSRVAVTTQHLLSIISLANTLMSMTNATFFIGDHMKKG	924	
DB	1062	FDLLLPHE--LRGHNCVRC-----SAFVSVDSTLLA-----TGD-----	1094	
QY	925	STPDLSKARGSPPTSSNIVQGIQKVAAPVVSARSADHSG	965	
DB	1095	---DNGBIR-----IWNVSGNELLHLHCAP--LSBEGAATHGG	1126	

RESULT 12

US-09-949-016-10066
; Sequence 10066, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10066
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10066

Query Match 2.1%; Score 163; DB 4; Length 1196;

Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;

QY 115 QFQFVGQREGRLICHGYPILVVDATSLVLSVLSKISPDWISSMISIRSHRTQ 174
DB 330 EYVLQQLQKQFKRIKSSDYDEALDEAMSISV--EMLREDIKYITDLSIL--OK 382
QY 175 DTVALSVTGILKVMIVTISDMDQTEPIFEESKPIYQNCQISFCACFTORSLLVC 234
DB 383 DVKPTKVLCL--WMETE--EVEDILQ--FVNKSL-FC 417
QY 235 SKYWRVFDAGDYSLLCSPSENGQTWTGDFVSDVKVIWTENGQSYI-- 282
DB 418 DR--NGKFR--YLLHDLQVFLTEKNCQSLQDLHKKIITQ 456
QY 283 -YKLPAISCLPASDS-----FRSDVGAVENTLPPVQHI-- 314
DB 457 RYQPHPTLSPQDDCMYWNFLAYHMASAKMKELCALMFSLDWIKAKTELVPALHIE 516
QY 315 -----LLDRKDKELLICPVTFYFGCREYFHKLIIQDSSGRINWNI-----SDT 361
DB 517 FVEYRHILDEK-----CAVSENF-----QEF--LSINGHLGRQPPNIVQLGCEPET 564
QY 362 AD-----KQGSBEGLAMTTSISLQEAFFDKLNPFCAGIIDQLSVIPNSNEPLKVTA 412
DB 565 SEVYQAKLQAKQEVNGLMYLEWINKNITN-----LSRLVVRP-----TDA 608
QY 413 VYIPANGRLVCGREDGS-IVIVPATOT-AIVQLLQGEHMLRGPPIHRLGHRNKVTCL 470
DB 609 VY-----H-----ACFSEDQRIASCAGADTLQVFAKTGKLE-----IKAHEDEVLC 654
QY 471 LYPHQVSARVDORYLSGGVDFSVIITWDFSGEMKHIFCVHGGEITQLLVPPENC--SA 527
DB 655 AF-----STDDRFATCSQDKVKIWNMTGLVHTYDEHSEV-----NCCHFTN 700
QY 528 RVQHICCSV-ASPHSVGLLSLREKKICMLASRLFPPIQVIKWRPDDYLIVVCGS-DGSVY 585
DB 701 SSHLLLATGSSDCFLKLDLQKCEKNTMFGHTNSVNHCRFSP-DDKLLASADGTLK 759
QY 586 VQMMDTGALDRCVMGITAVEILNACD--EAVPAADVLSHPANLVKQAMTRSLAALKNN 643
DB 760 LWDA-TSANERKSINKQF-FLNLEDPQEDMEVIVKCCSWADG-----ARIMVAAKN- 810
QY 644 AHKLOTLATNLASASDKGNLPKYSHNSLWQAIAKTNLTDPD----- 587
DB 811 ---KIFLWNTDSRSKRVADCRGHL-SWVHGYMFSPDGSFSLTSSDDQTIIRLWETKVCNKS 866

QY 688 -----IHVLFDFVEALI-----IQLLTEASRPNTALISPENLOK 722
DB 867 AVMLKQEVDFVQENEVWMLAVDHIRRLQLINGRTGQIDYLT--AQVSCCLSPHLQYI 924
QY 723 ASSSDKGGSF---LTGKRAAVLFO---QVKET---IKENIKHLLDDEEED-- 767
DB 925 AFG--DENGAIILELVNRR---IFQRFQHKKTVMHIQTADKTLISSDDAEIQVWN 979
QY 768 -----IMQRRESDPEYRSKSKPLTLLEYNLMTDTAKLFMSCLHAWGL---NEVLDE 818
DB 980 WQLDKCIPLRGHQETVKDFRLKNS--RLJNS-----PDGTVKVNIIITGNKEKDF 1029
QY 819 VCLDRGLMKLPKCTVSGLLSRGSHMSLMLPGVNPACKLGHGKTEVGRK----- 868
DB 1030 VC-----HOGTVL-----SCDISHDATKFSSTADKTAKIWS 1061
QY 869 ----LPASEGVGKGTGVSRVATTQHLLSIIISLANTLMSMTNATFTGDHMKKGPTRPPR 924
DB 1062 FLLLPFLHE--LRGHNGCVRC-----SAFVSDSTLLA-----TGD----- 1094
QY 925 STPDLSKARGSPPTSSNIVQGOIKQVAPVWSARSADHSG 965
DB 1095 ----DNGEIR-----IWNVSGELLHLCAP-LSEEGAATHGG 1126

RESULT 13

US-09-092-508-16
; Sequence 16, Application US/09092508
; Patent No. 6291643
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAP-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6291643west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,508
; FILING DATE: 05-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,807
; FILING DATE: 05-JUN-1997
; APPLICATION NUMBER: 60/055,258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kattelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-092-508-16


```
Db 574 SEVYQAKLQAKQEVNQMGLYLEWINKNITN-----LSRLVVRPH-----TDA 617
Qy 413 VYIPAGRLVCGREDGS-IVIVPATQF-AIVQLQGEHMLRRGWPHTLRGHRNKVTCL 470
Db 618 VY---H---ACFSEDGQRIASCAGADKTLQVFKATGEKLE-----IKAHEDEVLC 663
Qy 471 LYPHQVSARVDQRYLISGGVDFSVIIWDIFSGEMKHIFCVHGGEITQLVPPENC---SA 527
Db 664 AF-----STDDRFIATCSVDKKIKWNSMTGBLVHTYDEHSQV-----NCCHFTN 709
Qy 528 RVQRCICSV-ASDHSVGLLSLRKCKICMLASRLFPQIVKWRPDDYLWVGS-DGSVY 585
Db 710 SSHLLLATGSSDCFLKDLNOKCEKENTWFGHTNSVNHCRFP-DBKLLASCSAGTLK 768
Qy 586 VQWMDTGALDRCVMGITAVEILNACD--EAVPAADVSLSPAVNLKQAMTRRSALAKNM 643
Db 769 LWDA-TSANERKGINVKQF-FLNLEDPQEDMEVIVKCCSWADG-----ARIMVAAKN- 819
Qy 644 AHKILQTLATNLASASDKGNLPKYSHNSLWQAIKTNLTDPD----- 687
Db 820 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMPFGSGSSPLTSSDDQTIRLWETKVKCKNS 875
Qy 688 -----IHYLFDFVREALI-----IQLLTEASRPNTALISPENLQK 722
Db 876 AVMLKQEVDFVQFENEVWVLAVDHIRRLQINGRTGQIDYLTE--AQVSCCCLSPHLQYI 933
Qy 723 ASGSSDKGGSF-----LTGKRAAVLFQ---QVKET---IKENIKEHLDDDEEDEE----- 767
Db 934 AFG--DENGAIETELVNR--IFQSRFOHKKTVMHIQPTADEKTLISSDDAEIQVWN 988
Qy 768 -----IMQRRESDEPYESSKSLPTLLEYNLTWDTAKLFMSCLHAWGL---NEVLDE 818
Db 989 WOLDKCIFLKGHQTQVDFRLLKNS--RLLSWS-----FDGTVKVMNIITGNKEKDF 1038
Qy 819 VCLDRGLMKPCHTQSVGSLLSRGHMSLMLPGYNQPACKLSHGKTEVGRK----- 868
Db 1039 VC-----HOGTVL-----SCDISHDATKPSSTSDAKTKIWS 1070
Qy 869 ----LPASGVGKGTGVSRAVTTQHLLSIISLANTLMGMTNATFICDHMKKGTPTRPPR 924
Db 1071 FDLLLPHE--LRGHNGCYRC-----SAFSDVDSILLA-----TGD----- 1103
Qy 925 STPDLSKAGSPPTSSNIVGQGIQKQVAPVWSARSADHSG 965
Db 1104 ---DNGEIR-----IWNVSGELLHLCAP--LSBEGAATHGG 1135
```

RESULT 15

```
US-08-988-856B-4
; Sequence 4, Application US/0898856B
; Patent No. 6291642
; GENERAL INFORMATION:
; APPLICANT: Weinstein, Jasminder
; TITLE OF INVENTION: No. 6291642el Mammalian Cell Cycle Protein
; FILE REFERENCE: 06843.0026-04; A-283 D
; CURRENT APPLICATION NUMBER: US/08/988,856B
; CURRENT FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-988-856B-4
```

```
Query Match 1.9%; Score 150.5; DB 3; Length 499;
Best Local Similarity 19.5%; Pred. No. 6.1e-05;
Matches 86; Conservative 68; Mismatches 160; Indels 126; Gaps 15;
Qy 997 DLGLGDKFRPPLLEMLARWQDRCLVREAAQALLAEALRLRLEQAGKEAIDAWAPYLPQ 1056
Db 9 DLHSLQLDAPINAPPARWQKAKEAGAPSPMRA-ANRSHSAGRTFG----- 57
```

```
Qy 1057 YIDHVISPGVTSEAAQTITTAP--DASGPEAKVQEBEEDLVDDDITTCGLSSVPQMKKIS 1114
Db 58 -----RTPGKSSSKVQTTPTSPKPGDDRYIPIHRSAAQME-----VASFLLSKENQSENSQ 105
Qy 1115 TSVEERRKQATAIVLLGVIGABFGAIEPPKLLTRPRSSSQIPEGFG-----LTSGSNY 1169
Db 106 TPTKKEHQXAWALNLNGF-----DVEEAKILRLSGKQONAPEGYQNRILKVLYSOKATP 158
Qy 1170 SLARHTCKALTFLLLQPPSPKLPHPHSTIRRTAIDLIGRGFTVWEPTMDVSAVLMGLLELC 1229
Db 159 GSSRKTCRYI-----PSLPDR-----ILDAPEIR 182
Qy 1230 ADAEKQLANITWGLPLSPAADS-----ARSARHALSLIATARPPAFTTIAKEVHRHTAL 1284
Db 183 NDYYLNLVDWSSGNVLAVALDNSVYLWSASSGDIQLQMQEQGYISSVA-WIKEGNYL 241
Qy 1285 AANTOSQONMHTTTTLARAKGEILRVIEILIEKMTDVLVVEVMDIIMYCLEGSLVKKK 1344
Db 242 AVGTSSAE-----VOLWDV-----QQQK 259
Qy 1345 GLQECFPAICRFYVMSYVERNHRHIAVGARHGSVALYDIRTK--COTIHGKGPITAVAF 1402
Db 260 RLRRNMTSHSARVGSLSW--NSYILSSGSRSGHIIHHHDVRAEHVATLSGHSQEVCGLRW 317
Qy 1403 APDGRYLATYSNTDISHISFW 1422
Db 318 APDGRHLASCGN-DNLVNVW 336
```

Search completed: June 30, 2005, 17:25:20
Job time : 53 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 00:23:10 ; Search time 2091 Seconds
(without alignments)
12663.320 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcaggaaacagccttgt.....agcaccgcttcagtgtctaa 4473

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn19808:*
2: Geneseqn19908:*
3: Geneseqn20008:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4473	100.0	4473	12	ADQ60211 Human rab
2	4469.8	99.9	4548	4	ADL18130 Human TGF
3	4460.2	99.7	4517	6	ABV99403 Human NOV
4	3549.8	79.4	4497	4	ADL18131 Mouse GF-
5	3537.8	79.1	4494	4	ADL18129 Rat TGF-b
6	3497	78.2	6224	8	ABX70864 Novel hum
7	2807.4	62.8	3316	10	ADA53749 Human cod
8	569.2	12.7	834	5	ADM19384 Novel hum
9	557.6	12.5	630	5	ADM19354 Novel hum
10	474.6	10.6	2578	11	ADM02366 Human CDN
11	393	8.8	4578	13	ADS96553 Drosophil
12	393	8.8	5221	4	ABL19261 Drosophil
13	393	8.8	8411	4	ABL19260 Drosophil
14	393	8.8	13668	4	ABL06236 Drosophil
15	191.8	4.3	491	5	ADM19627 Novel hum
16	150.2	3.4	585	12	ACH74667 Human gen
17	147.2	3.3	227	12	ACH88367 Human gen
18	103.8	2.3	501	12	ACH75227 Human gen
19	103	2.3	356	3	AAC10429 Human sec
20	99	2.2	99	12	ACH88927 Human gen

21	88.4	2.0	555	12	ACH74845	Ach74845 Human gen
22	84	1.9	84	12	ACH88545	Ach88545 Human gen
23	69.2	1.5	3493	12	ADQ63583	Adq63583 Novel hum
24	57	1.3	65	6	ABN30858	Abn30858 Rat eplc
25	48.8	1.1	2000	8	ADA71938	Ada71938 Rice gene
26	48	1.1	2152	2	AAQ70732	Aaq70732 TATA-bind
27	48	1.1	2152	2	AAQ70732	Aaq70732 TATA-bind
28	48	1.1	2152	2	AAQ70732	Aaq70732 TATA-bind
29	48	1.1	3496	10	ADF81713	Adf81713 Leukaemia
30	46.4	1.0	2000	8	ADA71938	Ada71938 Rice gene
31	41.8	0.9	852	13	ADT44265	Adt44265 Bacteria
32	40.8	0.9	1159	6	ABQ68518	Abq68518 Listeria
33	40.8	0.9	1395	6	ABQ70546	Abq70546 Listeria
34	40.6	0.9	774	10	ADE28694	Ade28694 Human NOV
35	40.6	0.9	774	12	ADM93439	Adm93439 Human NOV
36	40.6	0.9	1110	12	ADQ85055	Adq85055 Human tum
37	40.6	0.9	1110	13	ADQ87397	Adq87397 Human tum
38	40.6	0.9	1110	13	ACN38886	Acn38886 Tumour-as
39	40.6	0.9	2862	13	ADQ85056	Adq85056 Human tum
40	40.6	0.9	2862	13	ADQ87398	Adq87398 Human tum
41	40.6	0.9	2862	13	ACN38887	Acn38887 Tumour-as
42	40.6	0.9	2863	4	AAH16061	Aah16061 Human CDN
43	40.6	0.9	3023	4	AAH14005	Aah14005 Human CDN
44	40	0.9	1296	5	AAS70486	Aas70486 DNA encod
45	39.6	0.9	514	12	ACH71629	Ach71629 Human gen

ALIGNMENTS

RESULT 1

ADQ60211
ID ADQ60211 standard; DNA; 4473 BP.

AC ADQ60211;

DT 07-OCT-2004 (first entry)

DE Human rabconnectin 3-binding protein coding sequence.

KW human; rabconnectin 3; GDP/GTP exchange reaction promotion protein;
KW binding protein; calcium dependent exocytosis; gene; db.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 1..4473

FT /*tag= a

FT /product= "Human rabconnectin 3 and GDP/GTP exchange
reaction promotion protein-binding protein"

XX JP2004201673-A.

XX PD 22-JUL-2004.

XX PF 13-AUG-2003; 2003JP-00207500.

XX PR 01-NOV-2002; 2002JP-00319521.

XX PA (EISA) EISAI CO LTD.

XX DR WPI; 2004-556370/54.

XX DR P-PSDB; ADQ60212.

XX PT Novel protein that couples with rabconnectin 3 and GDP/GTP exchange
PT reaction promotion protein, useful for elucidating chemical substance
PT that activates or inhibits calcium dependent exocytosis.

XX PS Claim 4; SEQ ID NO 1; 52pp; Japanese.

XX CC The invention comprises the amino acid and coding sequence of a protein
CC that couples with rabconnectin 3 and GDP/GTP exchange reaction promotion
CC protein. The DNA and protein sequences of the invention are useful for

CC analysing immune tissue for locating the protein of the invention or for
CC analysing the expression level of the protein of the invention. The DNA
CC and protein sequences of the invention are useful for elucidating
CC chemical substances that activate or inhibit calcium dependent
CC exocytosis. The present DNA sequence encodes the human rabconnectin 3 and
CC GTP/GTP exchange reaction promotion protein-binding protein of the
CC invention.

Sequence 4473 BP; 1226 A; 1024 C; 1047 G; 1176 T; 0 U; 0 Other;

Query Match 100.0%; Score 4473; DB 12; Length 4473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4473; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy	1	ATGCGAGGAAA	CAGCCCTGTGTTCTACCCATGTGTTCTTTGGGGTCGAAAAGGCCCAACAT	60
Db	1	ATGCGAGGAAA	CAGCCCTGTGTTCTACCCATGTGTTCTTTGGGGTCGAAAAGGCCCAACAT	60
Qy	61	TGCATCTCAGCGGTACTTTTAA	CAGATGATGGGGCCAGCATGTAACAGGATGTCACGAC	120
Db	61	TGCATCTCAGCGGTACTTTTAA	CAGATGATGGGGCCAGCATGTAACAGGATGTCACGAC	120
Qy	121	GGACAAATATGTCCTCGGGATCTTT	CAGTAGAACAATGCAAAATTAATCTCGAGCACTGTTG	180
Db	121	GGACAAATATGTCCTCGGGATCTTT	CAGTAGAACAATGCAAAATTAATCTCGAGCACTGTTG	180
Qy	181	TTTGGTCATACAGCATCAATCACTT	GTGTTGCTTAAAGCTTGCTCCAGGTGACAAACAG	240
Db	181	TTTGGTCATACAGCATCAATCACTT	GTGTTGCTTAAAGCTTGCTCCAGGTGACAAACAG	240
Qy	241	TATATTGTGAGTGCATCTGAAA	AGTGGAGATGTGCTCTGGGATGTGAGTGTGTCAGA	300
Db	241	TATATTGTGAGTGCATCTGAAA	AGTGGAGATGTGCTCTGGGATGTGAGTGTGTCAGA	300
Qy	301	TGTATTGAAATTTACAAAATTTAG	CTTTGTCACACATACTGGCATACAGTTCTACCAAGTTCTCT	360
Db	301	TGTATTGAAATTTACAAAATTTAG	CTTTGTCACACATACTGGCATACAGTTCTACCAAGTTCTCT	360
Qy	361	GTTTGGGAATCAGCGAAGAAGGCT	TTTTATGCCACGGACATTAACCTGAAATCCCTTGTT	420
Db	361	GTTTGGGAATCAGCGAAGAAGGCT	TTTTATGCCACGGACATTAACCTGAAATCCCTTGTT	420
Qy	421	GTGGATGCTACAGCCCTGGAAGTA	TTATACTCCTTAGTATCAAAAGATATCACAGACTGG	480
Db	421	GTGGATGCTACAGCCCTGGAAGTA	TTATACTCCTTAGTATCAAAAGATATCACAGACTGG	480
Qy	481	ATTAGCTCCATGATGATTAATTCG	ATCCACCGAACACAGTGTGATGACACTC	540
Db	481	ATTAGCTCCATGATGATTAATTCG	ATCCACCGAACACAGTGTGATGACACTC	540
Qy	541	TCGGTGACTGGCATCCTGAAGTCT	CGAATGTGTTTACCTCGGAAATAGTGACATGCAGGAT	600
Db	541	TCGGTGACTGGCATCCTGAAGTCT	CGAATGTGTTTACCTCGGAAATAGTGACATGCAGGAT	600
Qy	601	ACTGAGGCAATATTGAGAGGAAT	CCAAACCAATTTATGTGAGATTCGCAAGCATC	660
Db	601	ACTGAGGCAATATTGAGAGGAAT	CCAAACCAATTTATGTGAGATTCGCAAGCATC	660
Qy	661	TCCTTTTGTGCAATTTACACAA	AGGTCACTTTTGTGTTGCTTCGCAATATTCGAGGGTG	720
Db	661	TCCTTTTGTGCAATTTACACAA	AGGTCACTTTTGTGTTGCTTCGCAATATTCGAGGGTG	720
Qy	721	TTCGATGCCGGAGACTATTCTCT	TGTTGTTTCAGTCTCTAGTGAATAATGACAGCATGG	780
Db	721	TTCGATGCCGGAGACTATTCTCT	TGTTGTTTCAGTCTCTAGTGAATAATGACAGCATGG	780
Qy	781	ACCGGGGGGACATTTGCTCATC	AGATAAGTCATCTTTCGACAGAAAAATGGGCCAAGT	840
Db	781	ACCGGGGGGACATTTGCTCATC	AGATAAGTCATCTTTCGACAGAAAAATGGGCCAAGT	840
Qy	841	TATATTACAAACTACCTGCCAGT	TGCTCCAGTCTAGTGAATCATTTCCGAGTGATGTG	900
Db	841	TATATTACAAACTACCTGCCAGT	TGCTCCAGTCTAGTGAATCATTTCCGAGTGATGTG	900

Db 4141 GACATCGGACTGGAATAATGTCAGACAAATCCATGAGACAAAGGACCAATCACTGCGAGTG 4200
Qy 4201 GCTTTTGCTCTGATGGAAGATATCTTGCACCTACTCAAACTGACAGCCACATTTCT 4260
Db 4201 GCTTTTGCTCTGATGGAAGATATCTTGCACCTACTCAAACTGACAGCCACATTTCT 4260
Qy 4261 TTTTGGCAGATGAACAGTCACTGCTGGGAAGCATCGGCATCTGAACTCGGCACCTCAG 4320
Db 4261 TTTTGGCAGATGAACAGTCACTGCTGGGAAGCATCGGCATCTGAACTCGGCACCTCAG 4320
Qy 4321 CTGCGTGTCAATTAACCTACCAAGTGCCCTCTGTGAGCCCGCTGCCCGCTCCAC 4380
Db 4321 CTGCGTGTCAATTAACCTACCAAGTGCCCTCTGTGAGCCCGCTGCCCGCTCCAC 4380
Qy 4381 AATGCCCTCAAGCTGGCCCGCTCATCTGCACTTCCAAACCGCAACGTCATCTCATGGCC 4440
Db 4381 AATGCCCTCAAGCTGGCCCGCTCATCTGCACTTCCAAACCGCAACGTCATCTCATGGCC 4440
Qy 4441 CATGACGGGAAGGAGCACCGCTTTCATGGTCTAA 4473
Db 4441 CATGACGGGAAGGAGCACCGCTTTCATGGTCTAA 4473

RESULT 2
AAD18130
ID AAD18130 standard; cDNA; 4548 BP.
XX
AC AAD18130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human TGF-beta Resistance Associated Gene (TRAG) cDNA.
XX
KW Human; Transforming growth factor-beta; TGF-beta; cellular proliferation;
KW phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;
KW oncogenesis; gene mapping; transgenic animal; chromosome 18; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 76..4548
FT FT /*tag= a
FT FT /product= "Human TRAG protein"
XX
PN W0200166739-A1.
XX
PD 13-SEP-2001.
XX
PF 12-FEB-2001; 2001WO-US004475.
XX
PR 07-MAR-2000; 2000US-0187572P.
PR 08-MAR-2000; 2000US-0187848P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Thorgerirsson SS, Sanders S;
XX
DR WPI; 2001-589942/66.
DR P-PSDB; AAE10802.
XX
PT Novel isolated polypeptide for studying and modulating mechanisms
PT involved in cellular proliferation comprises transforming growth factor-
PT beta resistance associated polypeptide fragment.
XX
PS Claim 7; Page 56-61; 89pp; English.
XX
CC The present sequence is a cDNA encoding human transforming growth factor
CC (TGF)-beta resistance associated (TRAG) protein. Human TRAG gene is
CC located on chromosome 18. TRAG DNA or protein is useful for studying and
CC modulating mechanisms involved in cellular proliferation, and for
CC modulating cellular phenotype. It is also used for screening aggressive
CC metastasis in cancer cells, and for targets used in cancer therapy. TRAG
CC protein is useful for evaluating factors that interact with and/or

CC control TGF-beta signalling for understanding both cell proliferation
CC control and oncogenesis, and in protein-protein interaction cells. TRAG
CC DNA is useful as hybridisation probe, in chromosome and gene mapping, in
CC the generation of antisense RNA and DNA, for the preparation of TRAG
CC protein by recombinant techniques, and for generating either transgenic
CC or knockout animals which, in turn, are useful in the development and
CC screening of therapeutically useful reagents
XX
SQ Sequence 4548 BP; 1255 A; 1035 C; 1056 G; 1202 T; 0 U; 0 Other;
Query Match 99.9%; Score 4469.8; DB 4; Length 4548;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCAGGAAACAGCGCTTCTTCTACCCATTGTTCTTTGGGTCGAAAAGCGCCACACAT 60
Db 76 ATGGCAGGAAACAGCGCTTCTTCTACCCATTGTTCTTTGGGTCGAAAAGCGCCACACAT 135
Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTCAAGAC 120
Db 136 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTCAAGAC 195
Qy 121 GGCAAAATATGCTCTGGGATCTTTTACAGTAGATGGGGCCACGATCGTAAACAGGATGTCAAGAC 180
Db 196 GGCAAAATATGCTCTGGGATCTTTTACAGTAGATGGGGCCACGATCGTAAACAGGATGTCAAGAC 255
Qy 181 TTTGGTCATACAGCATCAATCACTTGTGTTGTCTTAAAGCTTGTGCTTCCAGTCAACAACAG 240
Db 256 TTTGGTCATACAGCATCAATCACTTGTGTTGTCTTAAAGCTTGTGCTTCCAGTCAACAACAG 315
Qy 241 TATATTGTGAGTGCACTCTGAAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGATGSCAGA 300
Db 316 TATATTGTGAGTGCACTCTGAAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGATGSCAGA 375
Qy 301 TGATTGAATTTACAAAATTAGCTTGGCAACATCTGGCATACAGTTCTTACAGTTCTCT 360
Db 376 TGATTGAATTTACAAAATTAGCTTGGCAACATCTGGCATACAGTTCTTACAGTTCTCT 435
Qy 361 GTTGGGAATCAGGAGAGAGAGGCTTTTATGCCAGGACATTAACCTCGAAATCCCTTGT 420
Db 436 GTTGGGAATCAGGAGAGAGAGGCTTTTATGCCAGGACATTAACCTCGAAATCCCTTGT 495
Qy 421 GTGGATGTCTACCGACCTTGAAGTATTTATCTCTTAGTATCAAAAGATATCAACAGACTGG 480
Db 496 GTGGATGTCTACCGACCTTGAAGTATTTATCTCTTAGTATCAAAAGATATCAACAGACTGG 555
Qy 481 ATTAGCTCCATGAGTATTTTCCATCCCGAACAACAAGAGGACACAGTGGTAGCACTC 540
Db 556 ATTAGCTCCATGAGTATTTTCCATCCCGAACAACAAGAGGACACAGTGGTAGCACTC 615
Qy 541 TCGGTGACTGGGCATCTGGAAGTCTGGATTGTTTACCTCGGAATAAGTGACATGCAAGGAT 600
Db 616 TCGGTGACTGGGCATCTGGAAGTCTGGATTGTTTACCTCGGAATAAGTGACATGCAAGGAT 675
Qy 601 ACTGAGCCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC 660
Db 676 ACTGAGCCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC 735
Qy 661 TCTTTTGTGCAATTTACAAAGGTCACCTTTGGTGTGTGTTTCCAAATATTGGAGGGTG 720
Db 736 TCTTTTGTGCAATTTACAAAGGTCACCTTTGGTGTGTGTTTCCAAATATTGGAGGGTG 795
Qy 721 TTGATGTCGGGAGCATTTCTTGTGTTGTTAGGTCCTAGTCAAAATGAGACAGATAGG 780
Db 796 TTGATGTCGGGAGCATTTCTTGTGTTGTTAGGTCCTAGTCAAAATGAGACAGATAGG 855
Qy 781 ACCGGGGGGACATTTCTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGCAAAAGT 840
Db 856 ACCGGGGGGACATTTCTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGCAAAAGT 915
Qy 841 TATATTACAAACTACCTGCGAGTTCGCTTCCAGCTAGTGAATTCATTCGCACTGATG 900
Db 916 TATATTACAAACTACCTGCGAGTTCGCTTCCAGCTAGTGAATTCATTCGCACTGATG 975

Db	3136	TTGGAGTGAAGAGCCGACAGGCGCTCTTGGCGGAAGTGAAGAGATTGAGCAG	3195	
Qy	3121	GCAGGAGGAAGAACCTATGATGCTGGGCTCTTACTTACCTCAGTACATAGACCAC	3180	
Db	3196	GCAGGAGGAAGAACCTATGATGCTGGGCTCTTACTTACCTCAGTACATAGACCAC	3255	
Qy	3181	GTCAATATCACTGGAGTCAATCAGAAAGCGGCGCAGACTATACACAGGCTCCTGATGCC	3240	
Db	3256	GTCAATATCACTGGAGTCAATCAGAAAGCGGCGCAGACTATACACAGGCTCCTGATGCC	3315	
Qy	3241	TCAGGSCCTGAAGCAAAAGTCAGAGGAGGAGCAGTACCTTGTTCACCATGACATCAC	3300	
Db	3316	TCAGGSCCTGAAGCAAAAGTCAGAGGAGGAGCAGTACCTTGTTCAGCATGACATCAC	3375	
Qy	3301	ACTGGTTGCTTATCAAGTGTCCCAAAATGAAGAAATTTCTACATCTTACGAGGAAGA	3360	
Db	3376	ACTGGTTGCTTATCAAGTGTCCCAAAATGAAGAAATTTCTACATCTTACGAGGAAGA	3435	
Qy	3361	CGGAAGCAAGCTACCGCTATTTTACTTGGAGTAATAGGAGCTGAATTTGGTGTCTGAA	3420	
Db	3436	CGGAAGCAAGCTACCGCTATTTTACTTGGAGTAATAGGAGCTGAATTTGGTGTCTGAA	3495	
Qy	3421	ATTGAACCTCCTAAACTATTGACCAAGCTCGAAGCTTAGCCAAATTTCTGAGGATTC	3480	
Db	3496	ATTGAACCTCCTAAACTATTGACCAAGCTCGAAGCTTAGCCAAATTTCTGAGGATTC	3555	
Qy	3481	GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATATTTGCAAGGCACGTGAC	3540	
Db	3556	GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATATTTGCAAGGCACGTGAC	3615	
Qy	3541	TTTCTTCTGCTACAGCTCCAGCCCAAACTTCTCCACACAGCACTATCCGAAGACA	3600	
Db	3616	TTTCTTCTGCTACAGCTCCAGCCCAAACTTCTCCACACAGCACTATCCGAAGACA	3675	
Qy	3601	GCCATTGATCTGATTCGAGCTGGTTCACCTGTTGGAGCTTACATGATGTCGCGT	3660	
Db	3676	GCCATTGATCTGATTCGAGCTGGTTCACCTGTTGGAGCTTACATGATGTCGCGT	3735	
Qy	3661	GTTCGATGGGCTTCTCGAACTTTTGGCGATGCGGAGAAACAACTTCCCAACATCACA	3720	
Db	3736	GTTCGATGGGCTTCTCGAACTTTTGGCGATGCGGAGAAACAACTTCCCAACATCACA	3795	
Qy	3721	ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCGGCTCTCGAGGCAATGCCCTCTCG	3780	
Db	3796	ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCGGCTCTCGAGGCAATGCCCTCTCG	3855	
Qy	3781	CTCATTTGCCACCGCCAGACCACCGCTTCTATCACCACCATAGCCAAAGAGGTACACAGA	3840	
Db	3856	CTCATTTGCCACCGCCAGACCACCGCTTCTATCACCACCATAGCCAAAGAGGTACACAGA	3915	
Qy	3841	CATACGGCTCTTGACAGCAATACCCAAATCAGCAGAGATATGACACAACTCTTGCA	3900	
Db	3916	CATACGGCTCTTGACAGCAATACCCAAATCAGCAGAGATATGACACAACTCTTGCA	3975	
Qy	3901	CGAGCTTAAGGGGAAATTTGAGAGTCATTGAAATTTCTATTGAAAGATGCCCAAGAT	3960	
Db	3976	CGAGCTTAAGGGGAAATTTGAGAGTCATTGAAATTTCTATTGAAAGATGCCCAAGAT	4035	
Qy	3961	GTGTGTGATCTTCTCGTGGAGGTTATGGACATCATTTATGTCCTTGAAGGATCTTTA	4020	
Db	4036	GTGTGTGATCTTCTCGTGGAGGTTATGGACATCATTTATGTCCTTGAAGGATCTTTA	4095	
Qy	4021	GTAAAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCAAGGTTCTACATGGTCAGC	4080	
Db	4096	GTAAAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCAAGGTTCTACATGGTCAGC	4155	
Qy	4081	TATTATGAGCGGAATCAGCAATAGCAATGTTGGAGCTCGCCATCGTTTCAGTGGCCCTGTAC	4140	
Db	4156	TATTATGAGCGGAATCAGCAATAGCAATGTTGGAGCTCGCCATCGTTTCAGTGGCCCTGTAC	4215	
Qy	4141	GACATCCGAGCTCGAAATGTGACAAATCCATGGACACAAAGGACCAATCACTGCGATG	4200	
Db	4275	GACATCCGAGCTCGAAATGTGACAAATCCATGGACACAAAGGACCAATCACTGCGATG	4275	
Qy	4201	GCTTTTGTCTCTGATGGAGATATCTGCCACCTACTCAAAACACTGACAGCCACATTTCT	4260	
Db	4276	GCTTTTGTCTCTGATGGAGATATCTGCCACCTACTCAAAACACTGACAGCCACATTTCT	4335	
Qy	4261	TTTTGGCAGATCAACACGTCCTGCTGGGAAGCATCGGCATGCTGAACTCGGCACTCTCAG	4320	
Db	4336	TTTTGGCAGATCAACACGTCCTGCTGGGAAGCATCGGCATGCTGAACTCGGCACTCTCAG	4395	
Qy	4321	CTGCGCTGATTTAAACCTTACAGGTGCCCCCTGTGCAAGCCGCGTCCCGGCTCCAC	4380	
Db	4396	CTGCGCTGATTTAAACCTTACAGGTGCCCCCTGTGCAAGCCGCGTCCCGGCTCCAC	4455	
Qy	4381	AATGGCTCAAGCTGGCGGCTCATCTGGACTTCCAAACCGCAAGTCTCATCTGCGCC	4440	
Db	4456	AATGGCTCAAGCTGGCGGCTCATCTGGACTTCCAAACCGCAAGTCTCATCTGCGCC	4515	
Qy	4441	CATGACGGGAAGGACACCGCTTCATGGTCTAA	4473	
Db	4516	CATGACGGGAAGGACACCGCTTCATGGTCTAA	4548	
RESULT 3				
ID	ABV99403	standard; DNA; 4517 BP.		
XX	ABV99403;			
XX	27-JAN-2003	(first entry)		
XX	Human NOV36a	coding sequence.		
XX	Human; anti-HIV; cytostatic; antidiabetic; antihypertensive; cachexia; AIDS;			
XX	antifungal; cardiac; haemostatic; neuroprotective; anorectic;			
XX	antitumor; immunosuppressive; osteopathic; antiparkinsonian; cancer;			
XX	antifertility; cerebroprotective; gene therapy; NOVX; fertility;			
XX	metabolic disorder; diabetes; obesity; infectious disease; anorexia;			
XX	neurodegenerative disease; Alzheimer's disease; Parkinson's disease;			
XX	immune disorder; haematopoietic disorder; cardiovascular disorder;			
XX	bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;			
XX	metabolic syndrome X; wasting disorder; cell differentiation;			
XX	Single nucleotide polymorphism; SNP; cell proliferation; haematopoiesis;			
XX	wound healing; angiogenesis; gene; ds.			
XX	Homo sapiens.			
XX	Key	Location/Qualifiers		
XX	variation	replace(91,G)		
XX	variation	/tag= a		
XX	variation	/standard_name= "Single nucleotide polymorphism"		
XX	variation	replace(230,C)		
XX	variation	/tag= b		
XX	variation	/standard_name= "Single nucleotide polymorphism"		
XX	variation	replace(4250,G)		
XX	variation	/tag= c		
XX	variation	/standard_name= "Single nucleotide polymorphism"		
XX	WO200272771-A2.			
XX	19-SEP-2002.			
XX	08-MAR-2002; 2002WO-US007288.			
XX	08-MAR-2001; 2001US-0274101P.			
XX	08-MAR-2001; 2001US-0274194P.			
XX	08-MAR-2001; 2001US-0274281P.			
XX	08-MAR-2001; 2001US-0274322P.			
XX	09-MAR-2001; 2001US-0274849P.			
XX	12-MAR-2001; 2001US-0275235P.			
XX	13-MAR-2001; 2001US-0275578P.			
XX	13-MAR-2001; 2001US-0275579P.			
XX	13-MAR-2001; 2001US-0275601P.			

PR 14-MAR-2001; 2001US-0276000P.
PR 16-MAR-2001; 2001US-0276776P.
PR 19-MAR-2001; 2001US-0276994P.
PR 20-MAR-2001; 2001US-0277239P.
PR 20-MAR-2001; 2001US-0277321P.
PR 20-MAR-2001; 2001US-0277327P.
PR 20-MAR-2001; 2001US-0277338P.
PR 21-MAR-2001; 2001US-0277791P.
PR 22-MAR-2001; 2001US-0277813P.
PR 23-MAR-2001; 2001US-0278152P.
PR 26-MAR-2001; 2001US-0278894P.
PR 27-MAR-2001; 2001US-0278999P.
PR 27-MAR-2001; 2001US-0279036P.
PR 28-MAR-2001; 2001US-0279344P.
PR 30-MAR-2001; 2001US-0279995P.
PR 30-MAR-2001; 2001US-0280233P.
PR 02-APR-2001; 2001US-0280802P.
PR 02-APR-2001; 2001US-0280822P.
PR 02-APR-2001; 2001US-0280900P.
PR 04-APR-2001; 2001US-0281194P.
PR 13-APR-2001; 2001US-0283675P.
PR 30-APR-2001; 2001US-0287424P.
PR 02-MAY-2001; 2001US-0288066P.
PR 03-MAY-2001; 2001US-0288342P.
PR 03-MAY-2001; 2001US-0288528P.
PR 15-MAY-2001; 2001US-0291190P.
PR 16-MAY-2001; 2001US-0291099P.
PR 16-MAY-2001; 2001US-0291240P.
PR 30-MAY-2001; 2001US-0294485P.
PR 31-MAY-2001; 2001US-0294889P.
PR 31-MAY-2001; 2001US-0294899P.
PR 18-JUN-2001; 2001US-0299027P.
PR 19-JUN-2001; 2001US-0299303P.
PR 19-JUN-2001; 2001US-0299310P.
PR 10-JUL-2001; 2001US-0304354P.
PR 31-JUL-2001; 2001US-0309198P.
PR 16-AUG-2001; 2001US-0312903P.
PR 10-SEP-2001; 2001US-0318462P.
PR 12-SEP-2001; 2001US-0318770P.
PR 27-SEP-2001; 2001US-0325430P.
PR 27-SEP-2001; 2001US-0325681P.
PR 19-OCT-2001; 2001US-0330380P.
PR 31-OCT-2001; 2001US-0335301P.
PR 14-NOV-2001; 2001US-0332172P.
PR 14-NOV-2001; 2001US-0332271P.
PR 14-NOV-2001; 2001US-0332272P.
PR 14-NOV-2001; 2001US-0333184P.
PR 14-NOV-2001; 2001US-0333272P.
PR 21-NOV-2001; 2001US-0332094P.
PR 03-DEC-2001; 2001US-0337426P.
PR 03-DEC-2001; 2001US-0338092P.
PR 04-DEC-2001; 2001US-0337185P.
PR 03-JAN-2002; 2002US-0345705P.
PR 08-MAR-2002; 2002US-00093463.

(CURA-) CURAGEN CORP.

XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
XX Baldog FL, Li L, Zerhuesen BD, Tchernev VT, Gangolli EA, Vernet CAM;
PI Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
PI Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;
PI Zhong M;

XX WPI; 2002-732824/79.
DR P-PSDB; ABP70125.

XX New NOVX polypeptides and polynucleotides, useful for preventing,
PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
PT disorders, and asthma.

XX Claim 16; Page 230-232; 619pp; English.

XX The present invention relates to new isolated proteins (NOVX) and their
CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
CC any number from 1 to 48. The NOVX proteins and coding sequences are
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, preferably a NOVX-associated disorder.
CC The NOVX coding sequences and proteins are useful for treating, diabetes,
CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
CC obesity, infectious disease, anorexia, cancer-associated cachexia,
CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
CC disease, immune disorders, hematopoietic disorders, cardiovascular
CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
CC disturbances associated with obesity, metabolic syndrome X or wasting
CC disorders associated with chronic diseases or various cancers. The NOVX
CC coding sequences and proteins may also be used as targets for the
CC identification of small molecules that modulate or inhibit e.g.
CC neurogenesis, cell differentiation, cell proliferation, hematopoiesis,
CC wound healing and angiogenesis, in gene therapy, in generation of
CC antibodies that bind immunospecifically to NOVX substances for use in
CC therapeutic or diagnostic methods

XX Sequence 4517 BP; 1236 A; 1038 C; 1059 G; 1184 T; 0 U; 0 Other;

Query Match 99.7%; Score 4460.2; DB 6; Length 4517;
Best Local Similarity 99.8%; Pred. NO. 0;
Matches 4465; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGGCAGGAACAGCCCTGTTCTACCCATTGTTCTTTGGGGTGGAAAGCGCCACACAT 60
DB 20 ATGGCAGGAACAGCCCTGTTCTACCCATTGTTCTTTGGGGTGGAAAGCGCCACACAT 79
QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCTTAACAGATCTACGAC 120
DB 80 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCTTAACAGATCTACGAC 139
QY 121 GGNCAATATGCTCTGGGATCTTTCAGTAGAAGTCTGCAATTAATCTTCGAGCCTGTTG 180
DB 140 GGNCAATATGCTCTGGGATCTTTCAGTAGAAGTCTGCAATTAATCTTCGAGCCTGTTG 199
QY 181 TTTGGTCATACAGCAATCAATCACTGTTGTTGTCTTAAAGCTTGTCTTCCAGTGCAACACAG 240
DB 200 TTTGGTCATACAGCAATCAATCACTGTTGTTGTCTTAAAGCTTGTCTTCCAGTGCAACACAG 259
QY 241 TATATTGTGAGTCATCTGAAAGTGGAGAGATGTCCTCTGGGATGTGATGTGACAGA 300
DB 260 TATATTGTGAGTCATCTGAAAGTGGGAGATGTCCTCTGGGATGTGATGTGACAGA 319
QY 301 TGTATTGAATTTACAAAATTAGCTTGCACACATCTGGCATACAGTTCTACCAAGTTCTCT 360
DB 320 TGTATTGAATTTACAAAATTAGCTTGCACACATCTGGCATACAGTTCTACCAAGTTCTCT 379
QY 361 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCCAGGACATTACCTGAAATCTCTGTT 420
DB 380 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCCAGGACATTACCTGAAATCTCTGTT 439
QY 421 GTGGATGTCACAGCCCTTGAAGTATTATATCTCTTAGTATCAAGATATACACAGACTGG 480
DB 440 GTGGATGTCACAGCCCTTGAAGTATTATATCTCTTAGTATCAAGATATACACAGACTGG 499
QY 481 ATTAGTCCATGATGATTTATTCGATCCACCGAACACAGAGGACACAGTGTGACACTC 540
DB 500 ATTAGTCCATGATGATTTATTCGATCCACCGAACACAGAGGACACAGTGTGACACTC 559
QY 541 TCGGTGACTGGCATCTGGAAGTCTGGATTGTTTACCTCGGAATAAGTCACATGCAAGGAT 600
DB 560 TCGGTGACTGGCATCTGGAAGTCTGGATTGTTTACCTCGGAATAAGTCACATGCAAGGAT 619
QY 601 ACTGAGCCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAAGCATC 660
DB 620 ACTGAGCCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAAGCATC 679
QY 661 TCTTTTGTGCAATTTACAAAGGTCATCTTTGTTGTTGTGTTTCCAAATATTTCGAGGGTG 720

Db 680 TCTTTTGTGCAATTTACAAAGGTCACATTTTGGTTGTGTTCCTCCAAATATTGGAGGTA 739
Qy 721 TTCCGATGCCGAGACATATTCCTTGTGTGTGTTCAGTCCCTAGTCAAAATGACAGACATGG 780
Db 740 TTCCGATGCCGAGACATATTCCTTGTGTGTGTTCAGTCCCTAGTCAAAATGACAGACATGG 799
Qy 781 ACCGGGGGACATTTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGCAAAAGT 840
Db 800 ACCGGGGGACATTTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGCAAAAGT 859
Qy 841 TATATTTACAAACTACTGCCAGTTCCTTCCAGCTAGTGAATTCATTCGCCAGTGAATGG 900
Db 860 TATATTTACAAACTACTGCCAGTTCCTTCCAGCTAGTGAATTCATTCGCCAGTGAATGG 919
Qy 901 GGGAAAGGAGTTGAAAATTTAAATTCCTCTGTACACATATCTCTTGGATCGAAAAGAT 960
Db 920 GGGAAAGGAGTTGAAAATTTAAATTCCTCTGTACACATATCTCTTGGATCGAAAAGAT 979
Qy 961 AAAGAGTTGCTAATTTGTCTCTCTGTACTCTCGGTTCCTTATGGATGCGAGAAATATTTTC 1020
Db 980 AAAGAGTTGCTAATTTGTCTCTCTGTACTCTCGGTTCCTTATGGATGCGAGAAATATTTTC 1039
Qy 1021 CATAAACTGTAAATTCAGGGTGATTCCTCTGGAAGGTTGAATATTTGGAAACATATCAGAC 1080
Db 1040 CATAAACTGTAAATTCAGGGTGATTCCTCTGGAAGGTTGAATATTTGGAAACATATCAGAC 1099
Qy 1081 ACAGCTGATAACAGGGAAGTGAAGAGGCTGGCAATGACAACATTCATATTAGTTTGCAG 1140
Db 1100 ACAGCTGATAACAGGGAAGTGAAGAGGCTGGCAATGACAACATTCATATTAGTTTGCAG 1159
Qy 1141 GAGGCAATTTGATAAATCTGATCTCTGTCTGTGAATATATAGATCAGCTGAGTGTGAAT 1200
Db 1160 GAGGCAATTTGATAAATCTGATCTCTGTCTGTGAATATATAGATCAGCTGAGTGTGAAT 1219
Qy 1201 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGCACATGACCGA 1260
Db 1220 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGCACATGACCGA 1279
Qy 1261 CTTGTTTGTGGTCTGTAAGATGGAAGCATAGTTATTTGTACCTGCCACAGACGGCCATA 1320
Db 1280 CTTGTTTGTGGTCTGTAAGATGGAAGCATAGTTATTTGTACCTGCCACAGACGGCCATA 1339
Qy 1321 GTACAGCTGTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACTCACAGAACATCTC 1380
Db 1340 GTACAGCTGTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACTCACAGAACATCTC 1399
Qy 1381 CGTGGTCATCGGAACAAAGTCACATGTTTGTCTATATCCTCATCAGGTCCTCAGTCCGGTAT 1440
Db 1400 CGTGGTCATCGGAACAAAGTCACATGTTTGTCTATATCCTCATCAGGTCCTCAGTCCGGTAT 1459
Qy 1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1500
Db 1460 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1519
Qy 1501 TCTGGAGAAATGAAACATATCTTCTGTGTTCACTGGTGGTGAATTAATCAACTTCTAGTT 1560
Db 1520 TCTGGAGAAATGAAACATATCTTCTGTGTTCACTGGTGGTGAATTAATCAACTTCTAGTT 1579
Qy 1561 CCACTGAAACCTGTAGTGAAGAGTACAGCACTGCAATCTGCTCTGTAGCCAGTGAACAC 1620
Db 1580 CCACTGAAACCTGTAGTGAAGAGTACAGCACTGCAATCTGCTCTGTAGCCAGTGAACAC 1639
Qy 1621 TCAGTAGGACTTCAAGTTTGCAGAGAAAATGCAATAATTTGGGCACTCTCGTCACTT 1680
Db 1640 TCAGTAGGACTTCAAGTTTGCAGAGAAAATGCAATAATTTGGGCACTCTCGTCACTT 1699
Qy 1681 TTTTCTTATCAAGTAATCAAAATGAGGCTTCTGTGATTAATGATTTGGGTTGGGTTTCA 1740
Db 1700 TTTTCTTATCAAGTAATCAAAATGAGGCTTCTGTGATTAATGATTTGGGTTGGGTTTCA 1759
Qy 1741 GATGGTTCTGTGACGTCTGGCAAAATGATATCTGGTGCATTTGGATCGTTGTGTGATGGGG 1800
Db 1760 GATGGTTCTGTGACGTCTGGCAAAATGATATCTGGTGCATTTGGATCGTTGTGTGATGGGG 1819

Qy 1801 ATAAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTCTGCTGCTGTGATTCA 1860
Db 1820 ATAAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTCTGCTGCTGTGATTCA 1879
Qy 1861 CTTAGTCAATCCAGCAGTCAACCTTAAACAAAGCTATGAACAGAGCTAGTCTTGTGCTCTTT 1920
Db 1880 CTTAGTCAATCCAGCAGTCAACCTTAAACAAAGCTATGAACAGAGCTAGTCTTGTGCTCTTT 1939
Qy 1921 AAAAATATGGCCCATCATTAAGCTACAAACCTTGCACCTAACTCTTGTGGCTTCTGAGGCA 1980
Db 1940 AAAAATATGGCCCATCATTAAGCTACAAACCTTGCACCTTAACTCTTGTGGCTTCTGAGGCA 1999
Qy 1981 TCTGACAAAGGGAAATTTTACATAATATCTCATAACTCCCTGATGGTTCAAAGCAATAAAG 2040
Db 2000 TCTGACAAAGGGAAATTTTACATAATATCTCATAACTCCCTGATGGTTCAGCAATAAAG 2059
Qy 2041 ACAAACTTAACAGACCCGACATACATGTGCTATTTCTTTTGAATGTGGAAGGTTGATTT 2100
Db 2060 ACAAACTTAACAGACCCGACATACATGTGCTATTTCTTTGATGTGGAAGGTTGATTT 2119
Qy 2101 CAACTCTGACTGAAGAGCCTTAGGCCGAATACCTCTTATTTTCCCGCAGAGAAATTTG 2160
Db 2120 CAACTCTGACTGAAGAGCCTTAGGCCGAATACCTCTTATTTTCCCGCAGAGAAATTTG 2179
Qy 2161 CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTAACTGGAATAACAGAGCA 2220
Db 2180 CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTAACTGGAATAACAGAGCA 2239
Qy 2221 GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACCTCTCTTGAT 2280
Db 2240 GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACCTCTCTTGAT 2299
Qy 2281 GATGAAGAGAGGATGAGGAGATTAATGAGGCGAGAGAGGAAGAAAGTCACTGGAATAT 2340
Db 2300 GATGAAGAGAGGATGAGGAGATTAATGAGGCGAGAGAGGAAGAAAGTCACTGGAATAT 2359
Qy 2341 CGGTCCAGCAAACTCAAGGCATTTGACCTATTAGATAATAATTTAACTATGACACTGCA 2400
Db 2360 CGGTCCAGCAAACTCAAGGCATTTGACCTATTAGATAATAATTTAACTATGACACTGCA 2419
Qy 2401 AAGCTGTTTATGCTCTGCTTCAAGCTGGGGTTGAAATGAAGTACTGGAATGAAGTTTGC 2460
Db 2420 AAGCTGTTTATGCTCTGCTTCAAGCTGGGGTTGAAATGAAGTACTGGAATGAAGTTTGC 2479
Qy 2461 CTGGAATCGCTTGGAAATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTTGTCAAGA 2520
Db 2480 CTGGAATCGCTTGGAAATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTTGTCAAGA 2539
Qy 2521 GGAGGCAATATGCTACTGATGCTGCGGGTTAATCAAGCTGCTTGTAAAACCTGTCAACAT 2580
Db 2540 GGAGGCAATATGCTACTGATGCTGCGGGTTAATCAAGCTGCTTGTAAAACCTGTCAACAT 2599
Qy 2581 GGGAAAACAGAAAGTAGGAAGAGCTGCGAGGCTGTGAGGAGTAGGAAGGAACTTAC 2640
Db 2600 GGGAAAACAGAAAGTAGGAAGAGCTGCGAGGCTGTGAGGAGTAGGAAGGAACTTAC 2659
Qy 2641 GGAGTGTCCGCTGCGCTCACACAGCATCTCTGCTCTATCATTTTCTTTGGCAATACT 2700
Db 2660 GGAGTGTCCGCTGCGCTCACACAGCATCTCTGCTCTATCATTTTCTTTGGCAATACT 2719
Qy 2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGAATCATATGAAGAGGGTCTTACGAG 2760
Db 2720 TTAATGAGTATGACCAATGCACTTTTATTTGGTGAATCATATGAAGAGGGTCTTACGAG 2779
Qy 2761 CCACTTAGACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
Db 2780 CCACTTAGACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2839
Qy 2821 AATATTTGTCAGGACAGATTTAAACAAAGTTGCTGACCTGTGCTGTTCCGCTCGGTCTGAT 2880
Db 2840 AATATTTGTCAGGACAGATTTAAACAAAGTTGCTGACCTGTGCTGTTTCCGCTCGGTCTGAT 2899

PT Novel isolated polypeptide for studying and modulating mechanisms
PT involved in cellular proliferation comprises transforming growth factor-
XX beta resistance associated polypeptide fragment.

Claim 7; Page 62-66; 89pp; English.

XX The present sequence is a cDNA encoding mouse transforming growth factor
CC (TGF)-beta resistance associated (TRAG) protein. Mouse TRAG gene is
CC located on chromosome 18. TRAG DNA or protein is useful for studying and
CC modulating mechanisms involved in cellular proliferation, and for
CC modulating cellular phenotype. It is also used for screening aggressive
CC metastasis in cancer cells, and for targets used in cancer therapy. TRAG
CC protein is useful for evaluating factors that interact with and/or
CC control TGF-beta signalling for understanding both cell proliferation
CC control and oncogenesis, and in protein-protein interaction cells. TRAG
CC DNA is useful as hybridisation probe, in chromosome and gene mapping, in
CC the generation of antisense RNA and DNA, for the preparation of TRAG
CC protein by recombinant techniques, and for generating either transgenic
CC or knockout animals which, in turn, are useful in the development and
CC screening of therapeutically useful reagents

XX Sequence 4497 BP; 1144 A; 1153 C; 1147 G; 1053 T; 0 U; 0 Other;

Query Match 79.4%; Score 3549.8; DB 4; Length 4497;
Best Local Similarity 87.3%; Pred. No. 0;
Matches 3903; Conservative 0; Mismatches 567; Indels 3; Gaps 1;

QY	1	ATGCGAGGAACAGCCCTGTCTTACCCATTGTTCTTTGGGGTGGAAAGCGGCCACACAT	60
DB	28	ATGCGAGGAACAGCCCTGTCTTACCCATTGTTCTTTGGGGTGGAAAGCGGCCACACAT	87
QY	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATCAGCAC	120
DB	88	TGCATTTCTGTCATCTGTTGACAGATGATGGGGCCACAAATGTAACTGCGATGCCACGAT	147
QY	121	GGACAAATATGCTCTGGGATCTTTCAGTAGAAGTGCATAATTAATCCTCGAGCAGCTGTG	180
DB	148	GGACAAATATGCTCTGGGATGTTTCGTAAGATGTTTCGTAAGATTAATCCCGACGACGTGA	207
QY	181	TTTGGTTCATACAGCATCAATCACTGTTGTTGTCTTAAAGCTTGCTTCCAGTGCAAAACAG	240
DB	208	TTTGGCCACACAGCATCCATCACTGTTGTTGTCAAAGCCTCGCTTCTGGGACAAAGCGG	267
QY	241	TATATTGTGAGTGCATCTGAAAGTGAGAGATGTGCTCTGGGATGTGAGTGTGATGGCAGA	300
DB	268	TACACTGTGAGCGCGTCTGCAACCGAGAGATGTGCTCTGGGATGTGAAACGATGGCAGA	327
QY	301	TGTATTGAATTTACAAAATTAGCTTGCACATACCTGGCATACAGTTCTACCAAGTTCTCT	360
DB	328	TGTATTGAATTTACCAAGTTAGCTTGCACACATCTGGCATACAGTTCTACCAAGTTCTCT	387
QY	361	GTTGGGAATCAGCGAGAAGGCTTTTATGCCACGACATTTACCTGAAATCCCTTGT	420
DB	388	GTTGGGAACAGCAAGAGGCGCCTCTCTGCCATGGACATTTACCTGAAATCCCTGTT	447
QY	421	GTGGATGCTACCGCTTGAAGTATTTATCTCTCTAGTATCAAAGATATCACAGACTGG	480
DB	448	GTGGATGCCACCGCTTGAAGTATTTATCTCTCTAGTATCAAAGATATCTCTCAGACTGG	507
QY	481	ATTAGCTCCATGATGATTTATTCGATCCCGCAACACAGAGGACACAGTGGTAGCACTC	540
DB	508	ATCAGCTCCATGATGATTCATCCACTCTCAGCGACACACAGAGGACACTGTGGTGGCGCTC	567
QY	541	TCGGTGATCTGGCATCTCTGAAAGTCTGGATTTGTTACCTCGGAAATAAGTGCATGCAAGAT	600
DB	568	TCTGTGACAGTATCTTGAAGGTGTGATTTGTCACTCTGAAATGAGTGGAAATGCAAGAT	627
QY	601	ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTTGTCAAGATTTGCCAAAGCATC	660
DB	628	ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTTGTCAAGATTTGCCAAAGCATC	687
QY	661	TCCTTTTGTGCAATTTACAAAGGTCACTTTTGTGTTGTTTCCAAATATTGAGGGTG	720

DB	698	TCCTTTTGTGCAATTCACACAGAGGTCACTTTTGGTCTGATGCTCTCAAAATATCTGGAGGTG	747
QY	721	TTGATGCGGAGAGACTATTCCTTGTGTGTTTCTGAGTCTCTAGTGAATAAGCAGACATCG	780
DB	748	TTTGTGCTGGCGACTACTCTCTGTGTGCTCAGTCTCTAGTGAATAAGCAGACATCG	807
QY	781	ACCGGGGGGACTTTGTCTCATATAAGTCACTATTGTCAGAGAAAATGGGCAAGT	840
DB	808	ACTGAGGGGACTTTGTGCTGTCGACAAAGTCACTATCTGACAGAAAATGGGCAAGT	867
QY	841	TATATTTTACAACTACCTGCCAGTTGCCCTCCAGCTAGTGTATTCCTGCGAGTGTG	900
DB	868	TACATCTACAACTCCCTGCCAGTTGCCCTCCAGCTAGTGTATTCATTCGCGAGCGAGTG	927
QY	901	GGGAGGCGAGTTGAAAATTTAATCTCTCTCTGTAACAATATCTCTTGGATCGAAGAT	960
DB	928	GGGAAAGCAGTGGAAAATTTGATCTCTCTGTCGAGCATAGCTCTTTGGATCAGAAAGAT	987
QY	961	AAAGAGTTGCTAATTTGCTCTCTCTGTTACTCGGTTCTCTATGGATGACAGAAATATTTTC	1020
DB	988	AAAGAGTTGCTAATTTGCTCTCTCTGTTACTCGGTTCTCTACGGATGCAAGGAATTTTG	1047
QY	1021	CATAAATCTTTAATTCAGGGTGATTTCTCTCGAAGGTTGAATATTTGGAAATATCAGAC	1080
DB	1048	CATAAGCTACTAATTCAGGGCGATTTCTCTGGAAGTTAAAATATCTGGAACATAGCAGAC	1107
QY	1081	ACAGCTGATAACAGGGAAGTGAAGGGCTGGCAATGACAACTCTATTTAGTTTGCAC	1140
DB	1108	ATAGCAGAGAAAACAGGAAGCCGATGAAGGGCTTAAAGATGACAACTTGTATTTAGTTGCAA	1167
QY	1141	GAGGCATTTGTATAAATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1200
DB	1168	GAGGCATTTGTACAGCTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1227
QY	1201	CCCAATAGTAAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGCAATGAGCA	1260
DB	1228	CCCAACAGCAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGCAATGAGCA	1287
QY	1261	CTTGTGTTGGTCTGTAAGATGGAAGTAACTGTAAGTAACTGTAAGTAACTGTAAGTAACTG	1320
DB	1288	CTTGTGTTGGTCTGTAAGATGGAAGTAACTGTAAGTAACTGTAAGTAACTGTAAGTAACTG	1347
QY	1321	GTACAGCTGTGCAAGGGGAAACACATGCTCAGAAAGGTTGGCCACCTCACAGAAACCTC	1380
DB	1348	GTTACGATTTGCAAGGAAACACATGCTCAGAAAGGTTGGCCACCTCACAGAAACCTC	1407
QY	1381	CGTGTGTCATCGGAAACAAAGTCACTGTTTGTATATCTCATCAGGTCTCAGCTCGGTAT	1440
DB	1408	CGTGGCCATCGGAAACAAAGTCACTGTTTGTATATCTCATCAGGTCTCAGCTCGGTAT	1467
QY	1441	GATCAAGATATCTGATATCTGAGGTGTGGATTTTTCATGATGATTTTGGGACATATTT	1500
DB	1468	GACCAAGATATCTGATATCTGAGGTGTGGATTTTTCATGATGATTTTGGGACATATTT	1527
QY	1501	TCGGAAGAAATCAAAATATCTGTTTCTGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1560
DB	1528	TCGGAAGAAATCAAAATATCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT	1587
QY	1561	CCACCTGAAAACTGTAGTGCAGAGTACAGCACTGCACTGCTGTGTAGCCAGTACAC	1620
DB	1588	CCCCCGGAAAACTGTAGTGCAGAGTTCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1647
QY	1621	TCAGTAGGACTTCTAAGTTTGGAGGAAAAATGCAATAATGTTGGCATCTCTGTCACCTT	1680
DB	1648	TCGTAGGCTCTCTAAGTTTGGAGGAAAAATGCAATAATGTTGGCATCTCTGTCACCTT	1707
QY	1681	TTTCTCTATTCAAGTAAATCAAAATGAGGCTTCTGATGATTTACCTGTTGGTGGGTGTTCA	1740
DB	1708	TTTCTCTATTCAAGTAAATCAAAATGAGGCTTCTGATGATTTACCTGTTGGTGGGTGTTCA	1767
QY	1741	GATGTTCTGTGTACCTCTGGCAAAATGGATACTGTGTGCAATTTGGATCTGTTGTGTGATGGG	1800
DB	1768	GACGGCTCTGTGTATGCTCTGGCAGATGGACACTGTGTGCGCTGTGATCTGTTGCAATGGGA	1827

1801 ATACAGCAGTTGAGATTCTAAAGCGTTGTGATGAAGCTGTTCTGCTGCTGTGTGATTC 1860
1828 ATACAGCAGTTGAGATACCTGAAAGCGTTGTGATGAAGCGTTCTGCGCTGTGAGACTCT 1887
1861 CTTAGTCATCCAGCAGTCAACCTTAAACAAAGCTATGACGAGCTAGTCTTCTGCTCTT 1920
1888 CTTAGTCATCCAGCAGTCAACCTTAAACAAAGCTATGACGAGCTAGTCTTCTGCGCTT 1947
1921 AAAAATATGCCCACATCAATAGCTACAAACCTTTGCAACTAACTCTTTGGCTCTTGAGGCA 1980
1948 AAAAATATGCCCACATCAATAGCTACAAACCTTTGCAAGCACTTTTGGCTCTTGAGGCC 2007
1981 TCTGACAAGGGAATTTACTAATATTTCTAATATTTCTAATATTTCTAATATTTCTAATATTT 2040
2008 TCTGACAAGGGAATTTACTAATATTTCTAATATTTCTAATATTTCTAATATTTCTAATATTT 2067
2041 ACAAACCTACAGACCGGACATACATGCTGCTATTTCTTTCATGTCGAGGCTTGTATTT 2100
2068 ACAAACCTACAGACCGGACATACATGCTGCTATTTCTTTCATGTCGAGGCTTGTATTT 2127
2101 CAATCTCTGACTGAAGAGCTCTAGGCGGAATATGCTCTTATTTTCCCGAGAGATTTG 2160
2128 CAATCTCTGACTGAGAGGCTCTAGGCGGAATATGCTGCTATTTTCCCGAGAGATTTG 2187
2161 CAAAAAGCATCTGGCAGTTGAGCAAAAGGGGCTCTTTTAACTGGAATAAGCAGCAGCA 2220
2188 CAGAAAGCATCTGGCAGTTGAGCAAAAGGGGCTCTTTTAACTGGAATAAGCAGCAGCG 2247
2221 GTTCTCTTCAACAGTGAAGAGCAATCAAGAGAGCAATCAAGAGAGCAATCAAGAGAGCA 2280
2248 GTTCTCTTCAACAGTGAAGAGCAATCAAGAGAGCAATCAAGAGAGCAATCAAGAGAGCA 2307
2281 GATGAAGAGAGCATGAGAGATTAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
2308 GAG 2367
2341 CGGTCCAGCAAAATCAAGAGCAATGAGCCTATTAAGAGATTAATTAATTAATTAATTAAT 2400
2368 CGGGCCAGCAAGTCCAGAGCACTGACCTACTAGATACAACTTACTATGAGATACGCA 2427
2401 AAGCTGTTATGTCCTGCTCTTCAAGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 2460
2428 AATATTATTCATGCTCTCTCAAGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 2487
2461 CTGAGTCGCTGGAATGCTGAAGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 2520
2488 CTGAGTCGCTGGAATGCTGAAGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 2547
2521 GGAGGCAATATGTCATGATGCTGCGGGTTAATCAAGCCTGCTGTTGAACTGTGACAT 2580
2548 GGAGGTCATATGTCGCTGATGCTTCCGGTTAATCAAGCCTGCTGGAAGCTGCTGAT 2607
2581 GGGAAACAGAGATGAGAGAGAGCTGCGAGCGTCTGAGGAGATGAGAGAGAGAGAGAGAG 2640
2608 GCGAAAGCAGAGATGAGAGAGAGCTGCCAGAGCGGAGCGGCTGAGGAGAGAGAGAGAG 2667
2641 GAGAGTCGCGTCCGCTGACCAACAGCAGCATCTCTGCTATCATTTCTTTGGCAATACT 2700
2668 ACAGTGTCTGAGCGGTGACCACTGAGCATCTGTTGTCATCATATATCCCTTGGGCAAT 2727
2701 TTAATGAGTATGACCAATGCAATTTTATTTGATGATCATATGAGAGAGGCTCTTACAGG 2760
2728 TTAATGAGTATGACCAATGCAATTTTATTTGATGATCATATGAGAGAGGCTCTTACAGG 2787
2761 CCACTAGACAG 2820
2788 CCGCTAG 2847
2821 AATATTGTCAG 2880
2848 AACATTGTCAG 2907

2881 GCTGATCAGTCTGGCTCTGACCCCTCTTCTGCTCTCTGCTCTTACATACCTGTTCTTAGTA 2940
2908 GCTGATCAGTCTGGCTCTGACCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2964
2941 AATGAAGTTGAGTCAAGTTAGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2965 AATGAAGTTGAGTCAAGTTAGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3024
3001 TTGGATAAATTTAGGCTCTCCCTTTCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3060
3025 CTGAGAGATTGAGGCTCTCTCTCTCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3084
3061 TTGAGGTGAG 3120
3085 TTGAGGTGAG 3144
3121 GCAGGAG 3180
3145 GCAGGAG 3204
3181 GTCATATCACTGAGTCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
3205 GTCATATCACTGAGTCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3264
3241 TCAGGAG 3300
3265 TCAGGAG 3324
3301 ACTGGTGTCTTCAAGTCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
3325 GCAGGTTGCTTCAAGTCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3384
3361 CGAG 3420
3385 AG 3444
3421 ATTGAACCTCTTAATATGAG 3480
3445 ATTGAACCTCTTAATATGAG 3504
3481 GGGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3540
3505 GGGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3564
3541 TTTCTTCTGCTACAGCTCCAG 3600
3565 TATCTTCTGCTACAGCTCCAG 3624
3601 GCCATTGATCTGATTTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3660
3625 GCCACTGATCTGATTTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3684
3661 GTTCTGATGGGCTTCTGAACTTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3720
3685 GTTCTGATGGGCTTCTGAACTTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3744
3721 ATGGGTTGCTCTGAGCCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3780
3745 ATGGGTTGCTCTGAGCCCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3804
3781 CTCTATGCGCAGCCAG 3840
3805 CTCTATGCGCAGCCAG 3864
3841 CATACGGCTTTGAG 3900
3865 CATACGGCTTTGAG 3924
3901 CGAGCTTAAAGGAG 3960
3925 AGGGCTTAAAGGAG 3984
3961 GTTGTGATCTTCTGAGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020

Db 688 TCTTTTGTGCAATTCACACAGAGGTGCTCTTGTGTTGATGCTCCAAGTACTGGAGGGTG 747
Qy 721 TTGATGCGCGGAGACTATTCCTTGTGTTGTTTCAAGTCTAGTGAATAATGACACAGATGG 780
Db 748 TTGATGCTGCGGACTACTCCTGCTGTTGTTTCAAGTCTAGTGAAGATGACACAGATGG 807
Qy 781 ACCGGGGGAGCTTTGTCTCATCAGATAAAGTCAATCTTTGGACAGAAATGCGCAAGT 840
Db 808 ACTGGAGGGGACTTTGTGCTGCGAGACAAGTCAATTTTGGACTGAATCGGGCAGAGT 867
Qy 841 TATATTTACAACTACCTGCGAGTTGCTTCCAGCTAGTGAATTCATTCGCGAGTGAATGG 900
Db 868 TACATTTACAACTCCCTGCGAGTTGCTTCCAGCTAGTGAATTCATTCGCGAGTGAATGG 927
Qy 901 GGGAGGCGAGTTGAAATTTAAATTCCTCTGTAACAATATTCCTTTGGATGCAAAAGAT 960
Db 928 GGGAAAGCAGTTGAAATCTGATTCCTCCGTCGAGCATAGACCTCTTGGATCAGAGGAT 987
Qy 961 AAGAGTTGCTAATTTGCTCCTGTTACTTGGTTCTTCTATGATGACAGAAATATTC 1020
Db 988 AGAGAGTTGTAATTTGCTCCTGTTACTCGGTTTCTTATGATGCAAGGAATATTTG 1047
Qy 1021 CATAAACTGTTAATTCAGGCTGATTCCTTGGAGGTTGAATATTTGGAAACATATCAGAC 1080
Db 1048 CATAAGCTACTAATTCAGGCTGATTCCTTGGAGGTTAAGTATTTGGAAACATAGCAGAC 1107
Qy 1081 ACAGCTGATAAACACAGGAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTGCA 1140
Db 1108 ATAGCAGACAAACAGGAAGCCAAATGAAGGCTTAAACACGAACTTGTATTAGTTGCA 1167
Qy 1141 GAGGCAATTCATAAATCCTGTTGCTGCTGCTGGAATTAATAGATCAGCTGAGTGAAT 1200
Db 1168 GATGCAATTTGACAACTGAAGCCCTGCTGCTGGAATTAATCAGCTGAGTGAAT 1227
Qy 1201 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGCACATGACGA 1260
Db 1228 CCAACAGCAACGAAACCACTTAAAGTAACTGCGAGTGTCTACATACAGCACACGGCGC 1287
Qy 1261 CTTGTTTGTGCTGTAAGTGAAGCATAGTTATTGTACTGCTGCCACACAGACGGCCATA 1320
Db 1288 CTTGTTTGGCGCGGGAAGACGAGCATCAATATCGTCCCTGCCACCCAGACGGCCATA 1347
Qy 1321 GTACAGCTGTTGACAGGGAACACATGCTCAGAAAGGTTGGCCACTCACAGAACATCT 1380
Db 1348 GTTCAGCTGCTGACGGGAGAACATGCTCAGACGAGGTTGGCCCCCGCACAGAACCCCTC 1407
Qy 1381 CGTGGTCACTCGGAACAAAGTCAATGTTGCTATATCCCTCATCAGTCTCAGCTCGGTAT 1440
Db 1408 CGTGGCCACCGGAACAAAGTCAATGTTGCTGTATCTCTCATCAGGTTCTCAGCTCGGTAT 1467
Qy 1441 GATCAAGATACCTGATATCTGGAGGTGTGGATTTTTCAGTCAATAATTTGGGACATATTT 1500
Db 1468 GATCAAGATACCTGATATCCGAGGTGTGGATTTTTCGTCATCAATTTGGGACATTTT 1527
Qy 1501 TCTGGGAAATGAAACATATCTTCTGTGTTTCAATGTTGGTGGAGATTAATCAACTTCTAGTT 1560
Db 1528 TCCGGGAAATGAAACATATCTTCTGTGTTTCAATGTTGGTGGAGATCAACAATCTCTGGTC 1587
Qy 1561 CCACCTGMAAAGTGTAGTGAAGTACAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Db 1588 CCGCCAGAAATCTGTAGTGAAGGTTCAACATCGCTGCTGTTCTGTGCTGCTGCTGCTGCTG 1647
Qy 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATATGATATGTTGGCATCTCGTCACTTT 1680
Db 1648 TCTGTAGGGCTGTAGTCTGCGAGAGAAATATGATATGCTGTGGGCTCTGTCACCTG 1707
Qy 1681 TTTCTTATTCAGTAATCAAAATGGAGCCCTTCTGATGATTAATCTGTTGGTGGGGTGTCA 1740
Db 1708 TTCCCTATTTCAGGTGATCAAGTGGAGGCTTCTGACGACTACCTGTTGGTGGGGTGCACG 1767
Qy 1741 GATGGTCTGTGTAGCTGCGGAAATGGATACCTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
Db 1768 GACGGCTCTGTGTGTGCTGCGAGATGGACATCGGTGCGCTGAGACCGCTGTGCAATGGGG 1827

Qy 1801 ATAAACGAGTTCAGATTTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTGATCA 1860
Db 1828 ATAAACGCGGTGAGATTTCTCAATGCTTGTGACGAAGCTGTCCCTGTCAGCAGTGGACTCA 1887
Qy 1861 CTTAGTCAATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGTGTGCTGCTGCTT 1920
Db 1888 CTCAGTCAATCCAGCAGTCAACCTTAAACAGCTATGACGAGGAGTGTGCTGCTGCTT 1947
Qy 1921 AAAAATATGCCCATCATTAAGCTTAAACCCCTTGAACCTTAACTCTTGGCTTCTGAGGCA 1980
Db 1948 AAAACATGCGCCACACAGCTTGAACCCCTTGAACCTTAACTCTTGGCTTCTGAGGCC 2007
Qy 1981 TCTGA CAAGGGAATTTA CTTAAATATCTCAAACTCTCCCTGATGTTCAAGCAATAAAG 2040
Db 2008 TCTGCA CAAGGGAATTTA CTTAAATATCTCAAACTCTCCCTGATGTTCAAGCAATAAAG 2067
Qy 2041 ACAAACTTAACAGACCGGACATACATGTGCTATTTCTTGAATGTGGAAGGTTGATTTAT 2100
Db 2068 ACAAACTTAACAGACCGGACATACATGTGCTATTTCTTGAATGTGGAAGGTTGATTTAT 2127
Qy 2101 CAACTCTGACTGAAGAGCCTCTAGGCCGAATACTGCTCTTATTTTCCCGCAGAGAAATTTG 2160
Db 2128 CAACTCTGACTGAAGAGCCTCTAGGCCGAATACTGCTCTTATTTTCCCGCAGAGAAATCTG 2187
Qy 2161 CAAAAAGCATCTGGCAGTTTCA CAAAAGGGGGCTCTTTTAACTGGAAGAACGAGCAGCA 2220
Db 2188 CAGAAAGCATCTGGCAGTTTCA CAAAAGGGGGCTCTTTTCTGACTGGAAGAACGAGCAGCA 2247
Qy 2221 GTTCTCTTCCACAGTGAAGAAAGCATCAAGAGAACTCAAGAAACACTCTCTTGTAT 2280
Db 2248 GTTCTTTTCCAGCAAGTGAAGAAACTATCAAGAGAACTAAGAGGAGCACTCTCTTGTAT 2307
Qy 2281 GATGAAGAGGAGATCAGGAGATAATGAGCAGAGAGGGAAGAAAGTGAATCTGGAATAT 2340
Db 2308 GAGGAGAGGAGCAGGAGAGGATGAGCAGGAGGAGGAGAAAGTGAATCTGAGTAC 2367
Qy 2341 CGTCCAGCAAACTCAAGGCCATTTGACCTATTAGAAATAAATTTAACTATGACACTGCA 2400
Db 2368 CGGGCCAGCAAGTCCAAACCACTTACCTACTAGATAACAACCTTACTATGAGATACGCA 2427
Qy 2401 AAGCTGTTATGCTCTGCTTCA CGCTGCGGTTTGAATGAAGTACTGGAATGAAGTTGTC 2460
Db 2428 AAATATTATCATGCTCTCTCCA CGCTGCGGTTTGAATGAAGTACTGGAATGAAGTTGTC 2487
Qy 2461 CTGGATCGCTTGGAAATGCTGAAA CCCCACCTGACCGTATCGTTTGGCTCTTGTCAAGA 2520
Db 2488 CTCGATCGCTTGGCAATGCTGAAA CCACTGCAAGTGTCTTTGGTCTCTCTATCCAGA 2547
Qy 2521 GGAGGCCATATGTCATGATGCTGCGGGTTATAATCAGCCTGCTTGTAAACTGTGCACAT 2580
Db 2548 GGAGGTCATATGCTTGTGATGCTTCTCTGTTATATCAGGCTGCTGGAAGCTACTGCAG 2607
Qy 2581 GGGAAAACAGATGAGGAAGAGCTGCCAGCGTCTGAGGAGTGAAGAAAGGAACTTAC 2640
Db 2608 GCGAAGCAGAAACAGAGCGAAGGGGCCAGCAA CCGAGAGCGTGAAGGAGGAGCTTAC 2667
Qy 2641 GGAGTGTCCGCTGCGCTCACCA CAGCATCTCTGCTCTATCATTTCTTTGGCAATACT 2700
Db 2668 ACAGTGTCCCGAGCGGTCAACCGCAACATCTGTTGTCTCATATATCTCTGGCAATACT 2727
Qy 2701 TTAATGAGTATGACCAATGCAACTTTTATTTGGTGTATCATATGAAGAGGGCTCTACAGG 2760
Db 2728 TTAATGAGCATGACCAATGCGAGCTTCA TTTGGAGATCATATGAAGAGGGGCCACAGG 2787
Qy 2761 CCACTTAGACCAAGCA CCCCAGACTTTCTAAGGCAAGGGGTTCCCTTCCAATCTTCAGT 2820
Db 2788 CCGCTTAGACCAAGCA CCCCAGACTTTCTAAGGCAAGGGATTTCCCGCCCGCAGCTCCAGT 2847
Qy 2821 AATATTGTGCAAGGACAGATTAAG CAAAGTGTGCTGACCTGCTGTTCCGCTCGGCTGAT 2880
Db 2848 AACATTGTGCAAGGACAGATTAAG CAAAGTGTGCTGCGCTGTCTCTCTGCTGCTGCTGCTG 2904

1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAAAATGCAATATGTTGGCATCTCGTACCTT 1680
1713 TCAGTAGGACTTCTAAGTTTGGAGAGAAAAATGCAATATGTTGGCATCTCGTACCTT 1772
1681 TTTCCATTCAAGTAATCAATGGAGCCCTTCTGATGATTAACCTGGTGGTGGGGTGTTC 1740
1773 TTTCCATTCAAGTAATCAATGGAGCCCTTCTGATGATTAACCTGGTGGTGGGGTGTTC 1832
1741 GATGGTCTGTGACGTCGTGGCAAAATGGATACCTGGTGCAATGGATCGTTGTGTGATGGG 1800
1833 GATGGTCTGTGACGTCGTGGCAAAATGGATACCTGGTGCAATGGATCGTTGTGTGATGGG 1892
1801 AATAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTCTGCTGTGATTTCA 1860
1893 AATAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTCTGCTGTGATTTCA 1952
1861 CTTAGTCAATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGTGTCTGTGCTCTT 1920
1953 CTTAGTCAATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGTGTCTGTGCTCTT 2012
1921 AAAAAATGSCCATCATAGCTACAAACCCCTTGAACCTAACCCTTTGGCTTCTGAGGCA 1980
2013 AAAAAATGSCCATCATAGCTACAAACCCCTTGAACCTAACCCTTTGGCTTCTGAGGCA 2072
1981 TCTGACAGGGAATTTACTTAATATTTCTAATCTCCCTGATGGTTCAAGCAATAAG 2040
2073 TCTGACAGGGAATTTACTTAATATTTCTAATCTCCCTGATGGTTCAAGCAATAAG 2132
2041 ACAAACCTAACAGACCCGACATACATGTGCTATTCTTTGATGGAAGCGTTGATTAAT 2100
2133 ACAAACCTAACAGACCCGACATACATGTGCTATTCTTTGATGGAAGCGTTGATTAAT 2192
2101 CAACTCCTGACTGAAGAGCCTTAGGCCGAATACCTGCTCTTATTTTCCAGAGAAATTTG 2160
2193 CAACTCCTGACTGAAGAGCCTTAGGCCGAATACCTGCTCTTATTTTCCAGAGAAATTTG 2252
2161 CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTAACTGGAAAACGAGCAGCA 2220
2253 CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTAACTGGAAAACGAGCAGCA 2312
2221 GTTCTCTTCCAAACAGTGAAGAAAACGATCAAGAGAAACATCAAGGAAACACCTCTTGAT 2280
2313 GTTCTCTTCCAAACAGTGAAGAAAACGATCAAGAGAAACATCAAGGAAACACCTCTTGAT 2372
2281 GATGAAGAGGAGTGAAGAGATAATGAAGCAGAGAGGGAAGAAAGTATCTGTAATAT 2340
2373 GATGAAGAGGAGTGAAGAGATAATGAAGCAGAGAGGGAAGAAAGTATCTGTAATAT 2432
2341 CGGTCCAGCAAAATCAAAAGCCATTGACCTATTAGAAATATAATTAATCTATGGACATGCA 2400
2433 CGGTCCAGCAAAATCAAAAGCCATTGACCTATTAGAAATATAATTAATCTATGGACATGCA 2492
2401 AAGCTGTTTATGTCCTGCTCTTACGCTGGGGTTTGAATGAAGTACTGGATGAAGTTTG 2460
2493 AAGCTGTTTATGTCCTGCTCTTACGCTGGGGTTTGAATGAAGTACTGGATGAAGTTTG 2552
2461 CTGGATGCGCTTGGAAATGCTGAACCCCACTGACCCGATATCGTTTGGCTCTTGTCAAGA 2520
2553 CTGGATGCGCTTGGAAATGCTGAACCCCACTGACCCGATATCGTTTGGCTCTTGTCAAGA 2612
2521 GGAGGCCATATGTCATGATGCTGCGGGTTAATCAAGCTGCTTGTAAACGTGCACAT 2580
2613 GGAGGCCATATGTCATGATGCTGCGGGTTAATCAAGCTGCTTGTAAACGTGCACAT 2672
2581 GGCAAAACAGAGTAGGAAGAGAGCTGCCAGCTCTGAGGGAGTAGGAAGGAACTTAC 2640
2673 GGCAAAACAGAGTAGGAAGAGAGCTGCCAGCTCTGAGGGAGTAGGAAGGAACTTAC 2732
2641 GGAAGTGTCCGTGCCGTGACCAACAGAGATCTCTGTCTATCATTTCTTTGGCAAAATCT 2700
2733 GGAAGTGTCCGTGCCGTGACCAACAGAGATCTCTGTCTATCATTTCTTTGGCAAAATCT 2792

QY 2701 TTAATGAGTATGACCAATGCAACTTTTATTTGGTGATCATATCAAGAAGGCTCTTACCAGG 2760
DB 2793 TTAATGAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAGGCTCTTACCAGG 2852
QY 2761 CCACCTAGACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
DB 2853 CCACCTAGACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2912
QY 2821 AATATTTGCAAGGACAGATTAACAAGTTTGTGTCACCTGTGCTGTTCCGCTCGGTCTGAT 2880
DB 2913 AATATTTGCAAGGACAGATTAACAAGTTTGTGTCACCTGTGCTGTTCCGCTCGGTCTGAT 2937
QY 2881 GCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTCTGCTTTACATACCTGTTCTTAGTA 2940
DB 2938 ----- 2937
QY 2941 AATGAAGTTGGAGTCAGTTAGTCTGCTATGCACTGTGTTATGCTGCCAGACCTTCTGGGA 3000
DB 2938 ---AAGGTTGGAGTCAGTTAGTCTGCTATGCACTGTGTTATGCTGCCAGACCTTCTGGGA 2993
QY 3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCGGAAGATGCGAAGATCGATGC 3060
DB 2994 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCGGAAGATGCGAAGATCGATGC 3053
QY 3061 TTGGAGGTGAGAGAACCGCAGACAGCCCTCTGCTTCTGGCGGAACTGAGAAGAAATGAGCAG 3120
DB 3054 TTGGAGGTGAGAGAACCGCAGACAGCCCTCTGCTTCTGGCGGAACTGAGAAGAAATGAGCAG 3113
QY 3121 GCAGCAGGAAGAAAGCCATTGATGCTGGGCTCTTACTTACTTACCTCAGTACATAGACCAAC 3180
DB 3114 GCAGCAGGAAGAAAGCCATTGATGCTGGGCTCTTACTTACTTACCTCAGTACATAGACCAAC 3173
QY 3181 GTCATATCACTCTGGAGTCACATCAGAACCGCGAGAGCTATCACCACCGCTCTCTGATGCC 3240
DB 3174 GTCATATCACTCTGGAGTCACATCAGAACCGCGAGAGCTATCACCACCGCTCTCTGATGCC 3233
QY 3241 TCAGGCGCTTGAAGCAAAAGTCCAGGAGGAAGAGCATGACCTTTGTTGAACGATGACATCAC 3300
DB 3234 TCAGGCGCTTGAAGCAAAAGTCCAGGAGGAAGAGCATGACCTTTGTTGAACGATGACATCAC 3293
QY 3301 ACTGTTGCTTATCAAGTGTGCC 3323
DB 3294 ACTGTTGCTTATCAAGTGTGCC 3316
RESULT 8
ADM19384
ID ADM19384 standard; cdNA; 834 BP.
XX
AC ADM19384;
XX
DT 20-MAY-2004 (first entry)
XX
DE Novel human channel/transporter gene #181.
XX
KW ds; gene; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neoptropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; gene therapy; channel/transporter protein;
KW rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KW cerebral ischemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.
OS Homo sapiens.
XX
XX WO200154472-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001307.
XX

PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217966P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232357P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-023935P.
PR 13-OCT-2000; 2000US-023937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476159/51.
DR P-PSDB; ADM19863.
XX
XX Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.

XX PS Claim 1; SEQ ID NO 191; 809pp; English.

XX CC The invention relates to an isolated nucleic acid molecule encoding a

CC channel/transporter protein or sequences at least 95% identical to a

CC these. The nucleic acids and proteins encoded by them are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,

CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used

CC in diagnosing a pathological condition or susceptibility to a

CC pathological condition. The antibodies to the proteins can also be used

CC in alleviating symptoms associated with the disorders and in diagnostic

CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays

CC (ELISA). Disorders which are diagnosed or treated include autoimmune

CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.

CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac

CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,

CC nervous system disorders e.g. Alzheimer's disease, infections caused by

CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.

CC The polypeptides can also be used to aid wound healing and epithelial

CC cell proliferation, to prevent skin aging due to sunburn, to maintain

CC organs before transplantation, for supporting cell culture of primary

CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can

CC also be used as a food additive or preservative to increase or decrease

CC storage capabilities. This sequence corresponds to a gene of the

CC invention.

XX SQ Sequence 834 BP; 211 A; 206 C; 208 G; 209 T; 0 U; 0 Other;

Query Match 12.7%; Score 569.2; DB 5; Length 834;

Best Local Similarity 98.6%; Pred. No. 2.3e-161;

Matches 574; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGCGAGGAACAGCCTGTGTTTACCCATGTTCTTTGGGGTCCGAAAGCGCCACACAT 60

DB 220 ATGCGAGGAACAGCCTGTGTTTACCCATGTTCTTTGGGGTCCGAAAGCGCCACACAT 279

QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAACAGATGTACGAC 120

DB 280 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAACAGATGTACGAC 339

QY 121 GGACAAATATGCTCTGGGATCTTTTCAGTAGAATCGAAATTAATCTCGAGCACTGTTG 180

DB 340 GGACAAATATGCTCTGGGATCTTTTCAGTAGAATCGAAATTAATCTCGAGCACTGTTG 399

QY 181 TTTGGTCATACAGATCAATCACTGTTGTTCTTAAAGCTTGCTTCCAGTGCAAAACAG 240

DB 400 TTTGGTCATACAGATCAATCACTGTTGTTCTTAAAGCTTGCTTCCAGTGCAAAACAG 459

QY 241 TATATTGTGAGTCATCTGAAAGTGAGAGATGTCCTCTGGGATGTGAGTGATGCAGA 300

DB 460 TATATTGTGAGTCATCTGAAAGTGAGAGATGTCCTCTGGGATGTGAGTGATGCAGA 519

QY 301 TGTTATGAATTTACAAATTTAGCTTGACACATCTGGCATACTGACATGTTCTACCACTGTTCT 360

DB 520 TGTTATGAATTTACAAATTTAGCTTGACACATCTGGCATACTGACATGTTCTACCACTGTTCT 579

QY 361 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCGACGACATTAACCTGGAATTCCTGTT 420

DB 580 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCGACGACATTAACCTGGAATTCCTGTT 639

QY 421 GTGGATGCTACGAGCCTTGAAGTATTATATCTCTTAGTATCAAGATATACCCAGACTGG 480

DB 640 GTGGATGCTACGAGCCTTGAAGTATTATATCTCTTAGTATCAAGATATACCCAGACTGG 699

QY 481 ATTAGTCTCATGAGTATTATTCGATCCACCGAACACAGAGGACACAGTGGTAGCACTC 540

DB 700 ATTAGTCTCATGAGTATTATTCGATCCACCGAACACAGAGGACACAGTGGTAGCACTC 759

QY 541 TCGGTGAGTGGCATCTCGAAGGTCTGGATTTGTTTACCTCGAA 582

DB 760 TCGGTGAGTGGCATCTCGAAGGTCTGGATTTGTTTACCTCGAA 801

RESULT 9

ADM19354

ID ADM19354 standard; cDNA, 630 BP.

XX AC ADM19354;

XX DT 20-MAY-2004 (first entry)

XX DE Novel human channel/transporter gene #151.

XX ds; gene; immunosuppressive; antiarthritic; antirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;

KW ophthalmological; gene therapy; channel/transporter protein;

KW rheumatoid arthritis; neoplasia; cardiac arrest; cerebrovascular disorder;

KW cerebral ischemia; angiogenesis; nervous system disorder;

KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;

KW epithelial cell proliferation; skin aging; sunburn; transplantation;

KW chemotaxis; food additive.

XX Homo sapiens.

XX WO200154472-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001307.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 13-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476159/51.
XX P-PSDB; ADM19833.
XX
XX Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; SEQ ID NO 161; 809pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a these. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a CC pathologic condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, CC nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. CC The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain CC organs before transplantation, for supporting cell culture of primary CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can CC also be used as a food additive or preservative to increase or decrease CC storage capabilities. This sequence corresponds to a gene of the CC invention.
XX
SQ Sequence 630 BP; 156 A; 163 C; 150 G; 156 T; 0 U; 5 Other;
Query Match 12.5%; Score 557.6; DB 5; Length 630;
Best Local Similarity 99.6%; Pred. No. 6.4e-158;
Matches 568; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 3904 GCTAAAGGGGAATTTTGAGAGTCATTGAAATCTTATTGAAAAGATGCCACAGATGTT 3963
DB 1 GCTAAAGGGGAATTTTGAGAGTCATTGAAATCTTATTGAAAAGATGCCACAGATGTT 60
QY 3964 GTGGATCTCTCGTGAGGTTATGCACATCATATGCTGCTTGAAGGATCTTAGTT 4023
DB 61 GTGGATCTCTCGTGAGGTTATGCACATCATATGCTGCTTGAAGGATCTTAGTT 120
QY 4024 AAAAAGAAAGGCTTTCAAGAATGTTTCCAGGCATCTGCAGGTTCTACATGGTCAGCTAT 4083
DB 121 AAAAAGAAAGGCTTTCAAGAATGTTTCCAGGCATCTGCAGGTTCTACATGGTCAGCTAT 180
QY 4084 TATGAGCGGAATCACAGAATAGCAGTTGGAGCTGCGCATGGTTTCAGTGGCCCTGTACGAC 4143

181 TATAGCGGATACAGATAGAGTTGGAGCTCGCATGGTTTCAGTGGCCCTGTACGAC 240
4144 ATCCGACTGGAAAATGTTCAGACAAATCCATGGGACAAAGGACCAATCACTGCACTGGCT 4203
241 ATCCGACTGGAAAATGTTCAGACAAATCCATGGGACAAAGGACCAATCACTGCACTGGCT 300
4204 TTTTCTCTGATGGAGATATCTTGGCAGCTTCTGACCTACTCAACACACTGACAGCCACATTTCTTTT 4263
301 TTTTCTCTGATGGAGATATCTTGGCAGCTTCTGACCTACTCAACACACTGACAGCCACATTTCTTTT 360
4264 TGGCAGATGAACACGTCCTGCTGGGAAGCATCGGCATGTCTGAATCTGGGACCATCTCAGCTG 4323
361 TGGCAGATGAACACGTCCTGCTGGGAAGCATCGGCATGTCTGAATCTGGGACCATCTCAGCTG 420
4324 CGCTGATTAACACCTTACAGGTGCCCCCTGTGACAGCCGCGTCCCCCGGCTCCCAAT 4383
421 CGCTGATTAACACCTTACAGGTGCCCCCTGTGACAGCCGCGTCCCCCGGCTCCCAAT 480
4384 GCCTTCAAGTGGCCCGGCTCATCTGGACTTCCAAACGCAAGTGCATCTCATGGGCCAT 4443
481 GCCTTCAAGTGGCCCGGCTCATCTGGACTTCCAAACGCAAGTGCATCTCATGGGCCAT 539
4444 GATCGGAAGGAGCACCGCTTCATGGTCTAA 4473
540 GATGGGAGGAGCACCGCTTCATGGTCTAA 569

RESULT 10

ADMO2366
ID ADMO2366 standard; cDNA; 2578 BP.

XX ADMO2366;

XX 20-MAY-2004 (first entry)

XX Human cDNA of the invention SEQ ID NO:1051.

XX ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.

XX Homo sapiens.

XX EP1347046-A1.

XX 24-SEP-2003.

XX 12-APR-2002; 2002EP-00008400.

XX 22-MAR-2002; 2002JP-00137785.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI; 2003-723558/69.
XX P-PSDB; ADM04809.

XX New polynucleotides and polypeptides are useful in gene therapy, for
XX developing a diagnostic marker or medicines for regulating their
XX expression and activity, or as a target of gene therapy.

XX Claim 1; SEQ ID NO 1051; 305pp; English.

XX The invention relates to a novel human polynucleotide and the encoded
XX polypeptide. A polynucleotide of the invention may have a use in gene
XX therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
XX as a primer for synthesizing the polynucleotide or as a probe for
XX detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are
XX useful in gene therapy, for developing a diagnostic marker or medicines
XX for regulating their expression and activity, or as a target of gene
XX therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides

CC are useful as pharmaceutical agents. The present sequence represents a
CC cDNA sequence of the invention.

XX SQ Sequence 2578 BP; 720 A; 531 C; 581 G; 746 T; 0 U; 0 Other;
Query Match 10.6%; Score 474.6; DB 11; Length 2578;
Best Local Similarity 55.1%; Pred. No. 3e-132;
Matches 999; Conservative 0; Mismatches 799; Indels 15; Gaps 3;

QY 29 TTGTTCTTTGGGGTCAAAGCGCCACACATTCATTCAGCTCAGCGGTACTTTTAAACAGATG 88
Db 276 TGGCACTCTCTGGGACAGAGGCCCTCCACAGCATCACTGCCATCATGATCACTGATG 335
QY 89 ATGGGCGCCACAGATCGTTAAACAGGATGTACAGCGGACAAATATGTCTCTGGGATCTTTTCA 148
Db 336 ACACGGAAACGATTTGTGACTGGAAGTCAAGAGGGTCAGCTCTGTCTCTGGAACTCTCTCAC 395
QY 149 TAGAACTGCAAAATTAATCCTCGAGCACTGTTGTTTGGTCATACAGCATCAATCACTTTGT 208
Db 396 ATGAACATAAAGATTTTCAGCGAAAGAACTCCTATTTTGGTTCATTCAGCTTCGGTAAACATGTT 455
QY 209 TGTCTAAAGCTTGTGTTCCAGTGACAAACAGTATATTGTGAGTGATCTGAAAGTGGAG 268
Db 456 TGGCAAGAGCAAGGGACTTCTCTAAACAGCCCTACATTTGTTAGTGTCTGTAANAATGGGG 515
QY 269 AGATGTGCTCTGGGATGTGAGTGGCAGATGTATTGAAATTTTACAAAATTTAGCTTGCA 328
Db 516 AGATGTGCTTTGGGAATGTCACCAATGGACATGGGTGGGAAGGCTACACTTCTTACA 575
QY 329 CACATACTGGCATACAGTTCTTACCAGTTCTCTGTTGGGAATCAGCGAGAAGAGCTTT 388
Db 576 GGCACACTGCAATCTGTTATTACCACTGCTCGTCCGGATGACAGAGAAGGCTGGCTTC 635
QY 389 TATGCCACGACATTTACCTGAAATCTTGTGTTGGATGCTACAGGCTTTGAAGTATTAT 448
Db 636 TTTGTTGTGGAGATATCAAGATGCTCTTATAATTTGATGCCAAAACCTTTTGGCTGTGTTTC 695
QY 449 ACTCTCTAGTATCAAAGATATCACCAGACTGATTTAGCTCCATGATGATATTATTTCCATCCC 508
Db 696 ACAGTTTATAGATCATCTCAGTTTCTGACTGGATCACTGCAATGTCATGTCATGTTCACTCCA 755
QY 509 ACCGAACACAGAGGACACAGTGGTAGCACTCTCGGTGATCGGCATCTCCGAAGGCTGGA 568
Db 756 TGAGAAATTCAGAAGATTCTCTCTGGTGGTATCAGTAGCTGCTGAGCTCAAAGTATGG 815
QY 569 TTTGTACCTCGGAAATAGTGAATGACAGGACTGAGCAATATTTTGAAGGAGATCCA 628
Db 816 ATCTTTCTCATCTATCAACAGCATTTCAAGAAAAGCAAGATGTCTATGAAAAGAAATCCA 875
QY 629 AACCAATTTTATGTCAGAAATTCGCAAGCATCTCTTTTGTGATTTTACACAAAGTCA 688
Db 876 AGTTCTTGAAGTCTTGAACCTCCAGACAAATTCGATTTTGCACATATATCTGAGAGACTTC 935
QY 689 TTTTGGTGTGTCTTCAAAATATTTGGAGGCTGTTCGATCCGAGACTATTTCTTGTGTT 748
Db 936 TATTGGTGTGTTTCTAAATGTTTGAAGGTTTATGATTATTGTTGATTTTCCCTCTGTC 995
QY 749 GTTCAGGTCCTAGTGAATAATGGAACATGGAACCGGGGGGAGACTTTGTCTCATCAGATA 808
Db 996 TGACGAAGTTAGTAGAAATGGGCAAGTCTTTTGTGTTGGAGAAGTGAATGCTGCTCACA 1055
QY 809 AAGTCATCATTTGGACAGAAAATGGCAAGTTATATTACAAACTTACCTGCCAGTTGCC 868
Db 1056 GAATCTCATCTGGACAGAAATGGTTCACAGTTTACATCTATCAGCTGCTGAACAGTGGGC 1115
QY 869 TTCCAGCTAGTGAATTCATTTCCGAGTGTGTTGGGAAGGAGTGTGAAAATTTTAAATTCCTC 928
Db 1116 TTTCAAAAAG-----CATATACCTGCTGATGGAAGAGTGTCTTAAAGAGACCATTTATC 1169
QY 929 CTGTACAAATATCTCTCTTGGATTCGAAAGATTAAGAGTTGCTTAATTTTCTCTCTCTGTTA 988
Db 1170 CTCATTTACTGTGCTCTACTTCTGTGTGAGAAAATTAAGGAACAGAGCCGCTCTCTGTTTA 1229

```
Qy 989 CTCGGTCTTCTATGATGACAGAGATATTTCCATAAAGTGTAAATTCAGGGTGAATCTT 1048
Db 1230 TGGCTACATGAATGAAGGAAGAGCGCTTTTATCAAGGTACTTTTCTCTGGAGAGTCT 1289
Qy 1049 CTGGAAGTTCGAATATTTGGAAACATATCAGACACAGCTG-----ATAAACAGGGAAGTG 1102
Db 1290 CAGGAGANATTAATTTGGGACATCCCTGATGTTCTGTATCCAAAGTTGATGGTCTC 1349
Qy 1103 AAGAAAGGCTGGCAATGACAACTCTATTAGTTTGCAGAGGCAATTTGATAAACTGAATC 1162
Db 1350 CTAGAGAGATACCAAGTAACGCCACCTGGACTCTTCAAGATAATTTTGTATAAGCATGATA 1409
Qy 1163 CTTGTCTCTGGAAATATATAGATCAGCTGAGTGTGATTCCTCAATAGTAATGAACCTCTTA 1222
Db 1410 CTATGTCAAAAGTAATTAATGACTATTTCTCTGGGCTTAAAGATGGGGCAGGAACCTGCTG 1469
Qy 1223 AAGTAACCTGCAAGTGTGTACATACAGACACATGACGACTTGTGTGGTCTGGAAGATG 1282
Db 1470 TAGTCACTTATCAGAGATATTTCAAGTCTTGTATAAAGTCTTGTATGCTGTGAAGATG 1529
Qy 1283 GAAGCATAGTTATTGTACCTGCCACACAGACAGCGCCATAGTACAGCTGTTGCAAGGGGAAC 1342
Db 1530 GGACAAATATCATTAACCAAGCTTTGAATGCTGCCAAAGCAAGACTTCTGGAAGTGGTT 1589
Qy 1343 ACATGCTCAGAGAGGTTGGCCACTCAGACAGACACTCCGTGTCTATCGGAACAAAGTCA 1402
Db 1590 CTTTAGT---AAAAGATTTCTCCCTCATAAAGTTCTTAAAGGCCACCAACCAAGTGTCA 1646
Qy 1403 CATGTTTGTATATCTCATCAGCTCTCAGCTCGGTATGATCAAGATACCTGATATCTG 1462
Db 1647 CTTCAATTAATCTACCAATGCTCTCTCTGGAATTAGACCAAGTTGATGTTGTCTG 1706
Qy 1463 GAGGTGTGGATTTTTCAGTCAATATTTGGACATATTTCTGGAGAAATGAACATATCT 1522
Db 1707 GGGACCTGGACTCATGTGATCTTGTGGATATCTTTACTGAAGAAATTTTGCATAAT 1766
Qy 1523 TCTGTGTTATGTTGGTGAATTAATCACTTCTAGTTTCACTGAAACCTGATGTCGAA 1582
Db 1767 TCTTTTGGAGCTGGTCCAGTAACAAGTCTTTTGTATGTCACCAAGAGATTTAAACTAA 1826
Qy 1583 GAGTACAGCACTGATCTGTCTAGCCAGTACCACTCAGTAGGACTTCTAAGTTTGC 1642
Db 1827 GGGGTGAGCAGATTAATTTGTGTGTGCGGTGACCACTCCGTGGCTCTCTTCAACCTTG 1886
Qy 1643 GAGAGAAAAATGATTAATTTGGCATCTCTGTCACCTTTTCTTATTTCAAGTAATCAAAAT 1702
Db 1887 AGGAAAGAGTTGCTCTGCATGCCCGGAAGCACCTTTTCTGTGAGGATGATAAAT 1946
Qy 1703 GGAGGCTTCTGATGATTAATCTGTGTGGTGGTGTGAGATGGTCTGTGTAAGTCTGGC 1762
Db 1947 GGCACCCGGTTGAGAAATTTTAAATTTGTGATGTGCAGATGACTCAGTTTATATCTGGG 2006
Qy 1763 AAATGGATCTGGTGCATGATGATGTTGTGTGATGGGATACAGCAGTTGAGATTTCTAA 1822
Db 2007 AAATTTGAAACAGGCACCTTTGGAAGACATGAGACAGGAGAGACGCAATTAATTTCTTA 2066
Qy 1823 ACGCTTGTGATGA 1835
Db 2067 ATTCTTGTGATGA 2079
```

RESULT 11

ADS96553

ID ADS96553 standard; cDNA; 4578 BP.

XX

AC ADS96553;

XX

DT 02-DEC-2004 (first entry)

XX

DE Drosophila melanogaster protein coding sequence, SEQ ID 174.

XX

KW Insecticide; Antiparasitic; Anthelmintic; gene; ds.

XX

Drosophila melanogaster.

XX

PN WO2004039999-A2.

XX

PD 13-MAY-2004.

XX

PF 08-AUG-2003; 2003WO-US024982.

XX

PR 30-OCT-2002; 2002US-0422377P.

XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX

PI Stam L, Kandar KP, Spana E, Bachmann J;

XX

DR WPI; 2004-376203/35.

XX

DR P-PSDB; ADS96554.

XX

Identifying a compound that inhibits the activity of a protein for Drosophila viability for use e.g., as insecticidal agent by expressing in a recombinant host a DNA molecule to produce a protein essential for Drosophila viability.

XX

PS Claim 1; SEQ ID NO 174; 57pp; English.

XX

The present invention relates to a method for identifying a compound that inhibits the activity of a protein essential for Drosophila viability. The method comprises: (a) expressing in a recombinant host a DNA sequence encoding a protein essential for Drosophila viability; (b) testing compounds suspected of having the ability to inhibit the activity of the protein expressed in (a); and identifying a compound tested in (b) that inhibits the activity of the protein. The method is useful in identifying a compound that inhibits the activity of a protein essential for Drosophila viability for use as insecticidal, ectoparasiticide, antiparasitic, anthelmintic or acaricidal agent. The present sequence is the DNA sequence for one such protein essential for Drosophila viability. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX

SQ Sequence 4578 BP; 1053 A; 1281 C; 1303 G; 941 T; 0 U; 0 Other;

Query Match

Best Local Similarity 8.8%; Score 393; DB 13; Length 4578;

Matches 2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;

Qy 1 ATGGCAGGAAACAGCGCTGTCTTACCCATTTGTTGGGTGCGAAAGCGCCACACAT 60

Db 1 ATGGTGAGTACCAATCTGGTGTACCCGTGGTGTGCTGTGGGTCCACCGCGCCACGCAC 60

Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAACAGGATGTACGAC 120

Db 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAACAGGATGTACGAC 120

Qy 121 GGACAAATATGTCTCGGGATCT---TTCAAGTAGAATCTGCAAAATTAATTCCTCAGACATG 177

Db 121 GGACAGATCTGCGCTTTGGCAAGTGGAGCCACACCCCTGGAAGATGTGCGCGGTGCTC 180

Qy 178 TTGTTTGGTTCATACAGCATCAATCACTTGTGTTGCTCTAAAGCTTGTGCTTCCAGTGACAA 237

Db 181 TTGTTGGGCATTCGGCGCCCGTTTGTGCTTAGTCGCTGCTCGCTTTCGCCGGAAC 240

Qy 238 CAGTATATTTGTGAGTGCATCTGAAAGTGGAGAGATGTGCTCTGGGATGTGATGATGGC 297

Db 241 AACTTCTCTGTCAGCTCGTCTGAGAACGAGAGATGTGCACCTGGGATCTTACGGACGGC 300

Qy 298 AGATGATTAATTAACAAAATAGCTTGCACACATCTGGGCATACAGTTCTACCAAGTTC 357

Db 301 AAGTGCATGGAGCGGTCAAGCTTCCACAGGTGCACACCAAAATTCAGAGTACCAACACC 360

Qy 358 TCTGTTGGGAATCAGCGAAGAGAGGCTTTTATGCCACGGACATTAACCTGAAATCCTT 417

Db 361 -----GCCAACAGCAGGAGATGTGCGCCTCTTTTTCATTTGGCTACTATGCTGAAATCATG 414

XX	Drosophila melanogaster genomic polynucleotide SEQ ID NO 9256.	
XX	Drosophila; developmental biology; cell signalling; insecticide;	
XX	pharmaceutical; gene; ds.	
XX	Drosophila melanogaster.	
XX	WO200171042-A2.	
XX	27-SEP-2001.	
XX	23-MAR-2001; 2001WO-US009231.	
XX	23-MAR-2000; 2000US-0191637P.	
XX	11-JUL-2000; 2000US-00614150.	
XX	(PEKE) PE CORP NY.	
XX	Venter JC, Adams M, Li PWD, Myers EW;	
XX	WPI; 2001-656860/75.	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more	
XX	genes from Drosophila and for elucidating cell signaling and cell-cell	
XX	interactions.	
XX	Claim 1; SEQ ID NO 9256; 21pp + Sequence Listing; English.	
XX	The invention relates to an isolated nucleic acid detection reagent	
XX	capable of detecting 1000 or more genes from Drosophila. The invention is	
XX	useful in developmental biology and in elucidating cell signalling and	
XX	cell-cell interactions in higher eukaryotes for the development of	
XX	insecticides, therapeutics and pharmaceutical drugs. The invention	
XX	discloses genomic DNA sequences (AB116176-AB130511), expressed DNA	
XX	sequences (AB101840-AB16175) and the encoded proteins (AB57737-	
XX	AB572072). The sequence data for this patent did not form part of the	
XX	printed specification, but was obtained in electronic format directly	
XX	from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX	Sequence 5221 BP; 1246 A; 1424 C; 1484 G; 1067 T; 0 U; 0 Other;	
XX	Query Match 8.8%; Score 393; DB 4; Length 5221;	
XX	Best Local Similarity 46.6%; Pred. No. 3.2e-107;	
XX	Matches 2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;	
Qy	1 ATGCACGAAACAGCCCTGTCTACCCATATGTTCTTTGGGTGGAAGCGCCACACAT 60	
Db	644 ATGGTAGTACCAATCTGGTGGTACCGTGGTGTCTGTGGGTCCACCGCGCCACGAC 703	
Qy	61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCCACGATCGTAAACAGGATGTACGAC 120	
Db	704 TGCATCTCAGCGGTCTCTCGGACGATCAGTTTACCTGGTTCACCGCGTCTACGAT 763	
Qy	121 GGACAAATATGTCTCTGGATCT---TTCAGTAGAATGCNAATATATCTCGAGCACTG 177	
Db	764 GGACAGATTCGCCCTTGGCAAGTGGAGCCACACCTGAAGATGTCCGCGCGGTGCCTC 823	
Qy	178 TTGTTTGGTTCATACAGCATCAATCACTTGTTTGTCTAAAGCTTGTGCTTCCAGTGACAA 237	
Db	824 TTGGTGGCCATTCGGCGCCGTTTTGTGCTTAGTGGCTGCCTCGCTTCTGCCGGAGAC 883	
Qy	238 CAGTATATTTGATGTGATCTGAAAGTGGAGAGATGTGCCCTCTGGGATGTGAGTGTGGC 297	
Db	884 AACCTTCCTGGTTCAGTCTGCTCGAAGACGGAGAGATGTGCACCTGGGATCTTACGGACGC 943	
Qy	298 AGATGTATTCGAATTTACAAATTAGTTTGCACATACTGSCATACAGTTCTACAGTTC 357	
Db	944 AAGTGCAATGGAGCGGCTCAGCTTCCACAGTGCACACCCNAATTCAGAGTACACACC 1003	
Qy	358 TCTGTTGGGAATCAGCGAGAGAGGAGCGCTTTTATGCCACGACATTAACCTCGAAATCTTT 417	
Db	1004 -----GCCAACACGAGGATGTGCGCTCTTTTGCATTTGGCTACTATGCTGAATCATG 1057	

1492 GACATATTTCTGAGAAATGAAACATATCTTCTGTGTTCTCATGCTGAGTACTCAA 1551
2126 GATCTGTACAGCGAAGCTTACTGCATAGGTTTTCGCTTTCATGCGGTGAGATCACAAA 2185
1552 CTTCTAGTTCACCTGAAAACCTGTAGTGCAGAGTACAGCACTGCATCTCTGTAGCC 1611
2186 CTGCTGTGCGCCAGAGAGCTGTAGTCCAGAAATCTGAAATGCAATATGTTTGGTGGCC 2245
1612 AGTGACCACTCAGTAGAGACTTCTAAGTTTGGAGAGAAAAATGCAATAATGTTGGCATCT 1671
2246 TCGGACCAATTCCTGTTACGCTAGTGTCCCTTCAGGAGCGCAAGTGTGTGACCCCTGGCAAGC 2305
1672 CGTCACCTTTTCTTATTCAGTAATCAAAATGAGGCTTCTGTGATGATTACCTGGTGGTG 1731
2306 AGACATCTATTCCTCTGTGTACCAATTAATGAGGCTCCCTGGACGACTTCTTAATCGTA 2365
1732 GGGTGTTCAGATGTTCTGTGTACGCTCTGGCAATGATATCTGTGCAATGGAATCGTGTGT 1791
2366 GGCTGTTCCGATGGCAGCGTGTATGTGTGGCAATGGAACGGGTACCTCGATCGAGTG 2425
1792 GTGATGGGATAACAGAGTTGAGATTCTAAACGCTTGTGATGAA----- 1836
2426 CTTTACCGGATGCTGGCAGAGAGTCTTTCGCTTTCGATGAAACAGGCGGAGGATGGC 2485
1837 -GCTGTTCTGCTGCTGTGATTCACTTAGTATCCAGCACTCAACTAAACAGCTAT 1895
2486 GGCTCGGGCGCGCGGTGTTCAATGGTGCATCTGCCAGGAGATGGGTATGGCCAAAT 2545
1896 GACGAGACGTAGTCTTCTGCTCTTAAATAATGCG-----CCATCAATAG 1941
2546 CCGGAGTGCATCTTCTTCCGTGGCTGAGTCCGGCAACATGATGCAATCCGTATGCT 2605
1942 CTACAAACCTTGCACTAACCTCTTGGCTTCTGAGGCATCTGACAGGGA----- 1994
2606 ACCCAGCGCGGATTTACGCAACTCCAGCACTTCAGSGTCAATAACAGGAATCTCGAC 2665
1995 --TTTACCTAAATTTCTCATACTCCTGATGTTCAAGCAATAAAGACAACTTAA 2052
2666 TTTTCTGATGAACACCGCAGCAATCTCTTTAGTATCCAGGCTCTCGTAAATCTCTAAG 2725
2053 GACCGGACATACATGTCTTATCTTTGATGTGAGCGTTGATTTCAACTCTCTGACT 2112
2726 GACGGGAGGACACATACCTTTCTTGAATCGAGGACTGATCTTGAATTTGCATAGC 2785
2113 GAAGAAGCTCTTAGCGCGAATACCTGCTTTATTTCCCGCAGAGAATTTGCAAAAGCATCT 2172
2786 GAGGAATACGACAGATGACACAGCTACCTTGGATCACTTGTGTGCTACCTGCAGAT 2845
2173 GGCAGTTCAGAAAGGGGCTCTTTTAACTGGAAACGAGCAGAGTCTCTTCCAA 2232
2846 CCAGAGATGGCAATCGATGCAATTTGGATGCTTAGCAAGAGATCGGTGACTTCTTCAAC 2295
2233 CAAGTGAAGAAACGATCAAGAGAAACATCAAGAAACACCTCTTGTATGATGAAGAGGAG 2292
2906 AAGTGAAGAAACAGGCGGTAGATGTGGAGAGATCTTGAAGCAAGAAACAAACCGGG 2965
2293 GATGAGGAGATAATGAGGCGAGAGAGGA----- 2321
2966 CTGGTCAGAAATTCAGGAGAAACCTGAGATTTGCGAGAGAAAGTACAGGCCAAGGTG 3025
2322 -----AGAAAGTATCTGAAATATCGGTCCAGCAATCAAAAGCCA 2361
3026 GAGAGCCTGCAAAAGGCGAGTAGAGCGCAGGAGCAGCAGGACCTGAAAGCAAAATC 3085
2362 TTGACCTATTAGAATAATTTAACTATGCACTGCAAGCTGTTATGTCCTGCCIT 2421
3086 GCTTCCAGATGAGGTTTACGATGTATGAGGATAGCCAGCTACTTCTTTCGCTGCTC 3145
2422 CAGCGCTGGGGTTGAATGAAGTACTGGATGAAGTTTGGCTGATCGCTTGGAAATGCTG 2481
3146 CATCTTGGGGCTAGATCCACATCTAGACAAAGATGTGGAGCGGCTTGGGCTGCTT 3205
2482 AAACCCACATGCAACCGTATCGTTTGGCTCTTGTCAAGAGAGGACCATATGTCATGATG 2541

3206 AGACCAATTTGTGCCCATCTCATAGGAGTGTCTTCAAGCGCGCTATATGTGCTGCTG 3265
2542 CTGCCGGGTTATAA-----TCAGCCTGCTTGTAAACTGTGCAC 2578
3266 CTGCCACATGGCAGAAACAACTATGCATACCAACCGGTATCCAGCTTCCCTCCAGTTCC 3325
2579 ATGGGAAAACAGAAAGTAGGAAG-----GAAGCTGCAGGCTCTGAGGGAGTA 2625
3326 AAGAAGCGCGCTGCCAGAGAGAGCTGCAGCGCTTGGAGCATCTCAGCGCTGTTTTCACA 3385
2626 GGAAGGGAACCTTACGAGGTGCCGTGCGTCACCAACAGCATCTCCTGTCTATCATTT 2685
3386 TCCCGTCTGCAITGGGAGCTCAGACTACGTTGACCACCAACCATATTTCTTGGCTGGT 3445
2686 TCTTTGGCAAAATCTTAAATGAGTATGACCAATGCAACTTTTATTTGTTGATATATGAAG 2745
3446 GCCATGTCANATAGCTGTCTGATGAGTGGCGCTCTCTTCTGCGGACGAGGAGAG 3505
2746 AAGGGTCTTACAGGCCACCTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTC 2805
3506 CACAAGAGCTGCAG--CGCTTGGCCCGCAGCGACTCCACATTTGAGCAACGAGGAGGA 3564
2806 CTTCCAACTTCCAGTAATATTTGTGCAAGGACAGATTAACAAAGTTGCTGCACCTGTCTT 2865
3565 GCGGAGGAGCTGATGGCCCAACCATCTCTCAAAATCAAGCACGCTGGAGCCTGCTTGC 3624
2866 TCCGCTCGGTCTGATGCTGATCACTCTGCTCTGACCCCTCTTCTGCTCTGCTTTACAT 2925
3625 TACCACCAATGCTTCTGTTGCCCGCAAAATTTGAGGCCCTGGAGCCGAGAAATTCAA 3684
2926 ACCTGTTTCTTAAATGAAGTTGGAGTTCAGTTAGCTGCTATGCACTGTGTTATGCTG 2985
3685 GCGTCCGCAAGTGGAAATGATGTTCAAGCGCTGGCAGCATCACTGCTATAG--ATCAGG 3742
2986 CCAGACTACTGGGATTTGATTAATTTAGCCCTCCCTTCTGAGATGCTGGCCCGAAGA 3845
3743 GAGGCGGCTCAGCAGATCTCTTTGGCGGAGCTTTACCCGATGGGCAAGAGGCTCGCAAG 3802
3046 TGGCAAGATCGATGCTTGGAGGTGAGAAAGCGCACAGCCCTGCTTCTGCGGGAACCTG 3105
3803 CAGTACTGAGAGCTGGGCAACATATCTGCCACTCTACGCAACACCGGAGCCATAGTG 3862
3106 AGAAGAATTTGAGCAGCAGCAGGAGAAAGCAATTTGATGCTGCGGCTCTTCTTACCT 3165
3863 GGAGCCCAACAGCAGTTGGCTTGAATGAGTCAAGCGGCAAGCGGAGAGCTGGCTCGGA 3922
3166 CAGTACATAGACCACTCATATCACTGAGTCAATCAGAGCCCGGAGCATATCAACC 3225
3923 TCGGCGGCAATGGAGGGGTGGCGTCCGAGTATCCGAGGCGGTGCTGCTGTTAGTGT 3982
3226 ACGGCTCTGATGCTCAGGCGCTGAAGCAAAAGTCCAGAGGAGAGCATGACCTTGT 3285
3983 AGCGGCGAGGTGGCAGCGTTCTTGGGGAGATGCCCCAAGATGAGACTACGAGAG 4042
3286 GAGCATGACATCACCCTGTTGCTTATCAAGTGTCCCAATTTGAAAAATTTCTTAC- 3344
4043 GAAGAGGAGGATTTATACGCAAAACCATCCAGCTTGTGCGAGCTGAAGCGCAAACT 4102
3345 -----ATCTTACGAGAAAGCAGGAGCAAGCTACCGTATTTGTTTACTTGGAGT 3395
4103 ACGGCGATTTATCTTACTGTTGTAAATAGGAGCAGAAATTTGGCCAGGATATATCCAGGAG 4162
3396 AATAGAGCTGAATTTG--GTGCTGAATTTGAACCTCTTAAATTTGACCAAGCTCGA 3453
4163 TCACCGAATCAACGTTGGAAGTATTTAGCATGGCCACGAGGAGCAATCTAACCTAGTGGAT 4222
3454 AGCTCTAGCAAAATTTCTTGGGATTTGCTGAGTGTGAGTTCGAATCTTACTGCTG 3513
4223 GCTGGCGGGAACGTCGAAATCTAGTGTGCTGGAGGCTTCCGGCATAGCCAACTCTA 4282
3514 GCGAGACATCTTGAAGGCACTGAGCTTTCTTCTGCTACAGCTTCCAAAGCCCAAACTT 3573

```
Db 4283 GCCGCTTACTTCAATGCCCCTGGCTCACTTGTACGGCCACCATCTCCAGCTG 4342
Qy 3574 CTTCCACACAGCACTATTCGAAGAAACAGCATTGATCTGATTTGGAGCTGGTTCAGTGT 3633
Db 4343 CCGCAATACACGCCACTGGCGCGGAGCCATCGATCTGTGGGTGCTGGATTACCGT 4402
Qy 3634 TGGGAGCCTTACATGATGATGTCGGCTGTTCTGATGGGGTCTTCGAACTTTGTGCCGAT 3693
Db 4403 TGGGAGCCTTACATGATGATGTCGAAGGTCTTCTTGGTTTGTGGAGATATGTCGGAG 4462
Qy 3694 GCCGAGAAACAACTTGCCAAACATCACAAATGGGTTGCCCTCTGAGCCAGCAGCTGACTCG 3753
Db 4463 GGC---AAGCACTGCCAATCTGAACACTCAAACTTCCACTAACCCCGCAAGCGGATGCC 4519
Qy 3754 GCCCGCTCTGCGAGGATGCCCTCTCGCTCATTTGCCACCGCCAGACACCGCGCTTCATC 3813
Db 4520 TGCGGAACCGCAGGCGATGCTCTGCGCTTGATAGCCACCGCTCGGCGCGCGGCTTCATC 4579
Qy 3814 ACCACCATAGCCAAAGGTTACACAGACATACGGCTTTCGACGAATACCCCAATCACAG 3873
Db 4580 ACTACGATGGCGCGGAGGTAGCCAGTACAATACGATGACGAGAACCGCGCAGTCCATC 4639
Qy 3874 CAGATATGACACAACTCTTTGCAC---GAGCTAAAGGGGAAATTTTGAGAGTCAIT 3930
Db 4640 AACACGCGCTACCCAGTCGGTTCTGCACAGGCCACGCGCAGATTTCTGCAGTGGTT 4699
Qy 3931 GAAATCTTTATGAAAGATGCCCAAGATGTTGTGGATCTTCTGTGAGGTTATGGAC 3990
Db 4700 GAGATGCTAATGCACAAATGTCAGTGGAGATCGCGGCTCTCTGGTGGAGTATGGAC 4759
Qy 3991 ATCATATGATGCTTCCCTTGAGGATCTTTAGTTAAAGAAAGGTTCTCAAGATGTTTC 4050
Db 4760 ATTGCCCTGCACTGCGTGGACGGAACGAGCTGAAGAACCGCGGCTTGGCGGAGTGTGT 4819
Qy 4051 CCAGCATCTGCAAGTCTCATGCTGCTGAGTATTTATGAGCGGAATCACGAATAGCAGT 4110
Db 4820 CTTGCCATCTGCAAGTTTAAACAGATCTGCCACTGGCCCAACACGCGCATCGCGTG 4879
Qy 4111 GGAGCTCGCATGGTTTCAGTGGCCCTGTACGACATCGGCACTGGAAATGTCAGACAATC 4170
Db 4880 GCGGCCAACAGTGGCAACCTGGCCATTTACGAGCTGGCCAGAACAGTGCACAGATGATC 4939
Qy 4171 CATGGACACAGGACCAATCACTGAGTGGCTTTTGTCTCTCATGGAAGATATCTTGC 4230
Db 4940 CCGGCCACACGATCGGATCACTGCTGGCTCTCTGCGCGATGGCAAGTACCTGCTG 4999
Qy 4231 ACCTACTCAAACTGACAGCCACATTTCTTTTGGCAGATGAACAGTCACTGCTGGGA 4290
Db 5000 TCGTACTTTGGCAGAGAACCGCTCTCTCTTGGCAGACCTCAACGGGGATGTTCCGA 5059
Qy 4291 AGCATGGCATGCTGAACCTCGGCACTCAGCTGGCTGGCTGCAATTAACCTTACAGGTGCC 4350
Db 5060 CTGGGC-----CAGTCGACAGCGGCTGTACGAAGGGCTACTCTACGGCA 5104
Qy 4351 CTTGTGACCGCGGCTCCCGGCTCCCAATGCCCTCAAGCTGGCCCGGCTCATCTGG 4410
Db 5105 CCCATTCGGATGTGTGC-----GCCTGAATCCGATGCGCTTGGCAGGCTGTGTGG 5158
Qy 4411 ACTTCCAAACGCAACGTCATCTCATGGCCCATGACGGGAAGGAGCAGCCTTTCATGGTC 4470
Db 5159 ATCAAGATGACCGGTGACCTCATGCTGGCCGACGGCTCCGAGAGCGGATTCATGTC 5218
Qy 4471 TAA 4473
Db 5219 TAA 5221
```

RESULT 13
ABLI19260
ID ABL19260 standard; DNA; 8411 BP.
XX
AC ABL19260;
XX

```
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9253.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WP1; 2001-656860/75.
XX
DR New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX
PS Claim 1; SEQ ID NO 9253; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
sequences (ABLI01840-ABLI16175) and the encoded proteins (ABBS5773-
ABBS72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8411 BP; 2232 A; 2053 C; 2136 G; 1990 T; 0 U; 0 Other;

Query Match 8.8%; Score 393; DB 4; Length 8411;
Best Local Similarity 46.6%; Pred. No. 4.4e-107;
Matches 2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;

Qy 1 ATGGCAGGAAACAGCCTTGTCTTACCATTTGTTGGGTGCGAAAGCGCCACACAT 60
Db 2834 ATGGTGAGTACCAATCTGGTGTACCCGTGGTCTGTGGGTCCCAACCGCCACGAC 2893
Qy 61 TGCATCTCAGCGTACTTTTACAGATGATGGGCCACGATGTAACAGATGTCAAGAT 120
Db 2894 TGCATCTCAGCGTGTCTCTCGACGATCAGTTTACCTGTCCCGGCTGTACGAT 2953
Qy 121 GGACAAATATGCTCTGGGATCT---TTCAGTAGAACTGCAAAATTAATCTCGAGACTG 177
Db 2954 GGACAGATCTGCTTTGGCAAGTGGAGCCACCAACCTGAAATGTGCGCGGTCCTC 3013
Qy 178 TTGTTTGGTCATACAGCATCAATCACTTGTGTTTGTCTAAAGCTTGTCTTCCAGTGA 237
Db 3014 TTGTTGGCCATTCGGCGCCCGTTTGTGCTTAGTGCCTGTCTCGCGGAGAAC 3073
Qy 238 CAGTATATTTGTGATGTCATCTGAAAGTGGAGATGTCCTCTGGGATGTGATGGC 297
Db 3074 AACTTCTGCTCAGCTCGTCTGAGAACGGAGATGTGCACCTGGGATCTTACGGACGGC 3133
Qy 298 AGATGATTTGAATTTTACAAATTTAGCTTGCACACATCTGGCATACAGTTCTACCA 357
Db 3134 AGTGCATGAGCGGCTCAGCTTCCACAGGTGCACACCCAAATTCAGAGTACCACAC 3193
Qy 358 TCTGTTGGGAATCAGCGAAGAGGCTTTTATGCCAGGACATTTACCTGGAATCTTT 417
```

Db 3194 -----CCCAACAGCGAGGATGTGCGCCTCTTTTGCATTGGCTACTATGCTGAAATCATG 3247
Qy 418 GTTGTGATGCTACCGACCTTGAAGTATATPACTCTCTAGTATCAAAAGATATCACCCAGAC 477
Db 3248 GTCATGATCCGTTTCAGCTAGAGGTGATCTACGTGCTCAGCTCAAAAGGTCAAAACCGGAT 3307
Qy 478 TGGATTAGCTCCATGAGTATTTATTCGATCCACCGAACAACAGAGGACAGTGGTAGCA 537
Db 3308 TGGATATCCGCTATTTACAGTACTGCGCCCATGCGTGCAGAGGATGACGTAGTGTGGCC 3367
Qy 538 CTCTCGGTGATGTCATCTCTGAAGGTCTGGATTGTTACCTCGGAAATAAGTGACATCGAG 597
Db 3368 ATCACCACACCGGAACCGTGAAGTCTGACCTCTGAC-----GGGTACAGAGACAAA 3421
Qy 598 GATATGAGCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTGAGATTTGCCAAAGC 657
Db 3422 CAGCGGAGGCCATTTACGAGAACGAATCCAAGGAGATCCGCTGCCTGAATGCCATCACC 3481
Qy 658 ATCTCTTTTGTGCTATTTACAAAGGTCTCTTTTGGTGTGTTCCAAATATTTGGAGG 717
Db 3482 ATGAATTTGCTGCGCCAGAACCCAGCGCACAGTGTCTTTAGTTGTACCAATATCTGGCAG 3541
Qy 718 GTGTTCGATCCGAGAGACTATCTCTGTGTGTTTCAGGTCTCTAGTGAATAATGGACAGACA 777
Db 3542 ATCTACGATGCGCGGACTTCACTGTACTCTGCTCCGTAAATGCTCCCGCAAGGACGT 3601
Qy 778 TGGACCGGGGGGACTTTGTCTCATCAGATAAAGTCAATATTTGGACAGAAAAATGGGCAA 837
Db 3602 TGGCAGGGCGGACTTTCATCACCTCCGACCGGGTAATGTATGGACAGATGAGGGAAG 3661
Qy 838 AGTTATATTTACAACTACCTGCGAGTTGCTTCCAGCTAGTATTCATTTCCGAGTGTAT 897
Db 3662 GGCTACTTTGTACAGCTGCGCGCAAACTGCAATTTCCGACAAAGGAGTTCCACTCGAAG 3721
Qy 898 GTGGGAAGGAGTGAATAAT-----TAATCTCTCTGTACAAACATATCTCTTGGATC 952
Db 3722 AGTGTGTGAGGATGACCTTACCTCTACTAGTGTGCAACGCGCGGTGATTAAGGTG 3781
Qy 953 GAAAGATAAAGAGTTGTCTAATTTGTCTCTCTGTACTCGGTTCTTCTATGGATGCAAG 1012
Db 3782 CTGTGTCGCCCCCAGCCATGAAGTTGCTTTCAGGCGCGGTGCGCCAGCAATCTACTG 3841
Qy 1013 AATATTTCCATAAATCTGTTAATTCAGGGTGAATTTCTTGAAGGTTGAATATTTGGACA 1072
Db 3842 CGCGGAGACTCCGAGGGATACATCTCTGTCTGGAATGTGCGGAGGTGCGGTGTAGACAAT 3901
Qy 1073 TATCAGACACAGCTGATAAAGAGGAGTGAAGAGGCTGGCAATGACAACT-TCATTT 1131
Db 3902 ATTAGCATCTGCAGGCCAAACAGATCCCGCGGACCCCTCAAGCGCACGTGTGCACT 3961
Qy 1132 AGTTTGCAGAGGCATTGTGATAAATCTGAATCTTGTCTCTGCTGGAATTTATAGATCAGCTG 1191
Db 3962 TCGTTGTTCAGGCGGTGTCCTAATAATGGAATCCACACAGTGGGCATCTCTGGATCAGC-- 4019
Qy 1192 AGTGTGATTTCCAAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTAATACCGACA 1251
Db 4020 -----TCTCCCGCATACAGAAATCGCGGTGAAGTCTCACCTCGAGCATCTATCTGCCACAG 4075
Qy 1252 CATGGACGACTGTTTGTGTGTGAGATGGAAGCATAGTTATTTGACCTGCGCACACAG 1311
Db 4076 CAGATCGCTGTGTATCGAGCGGAGATGGGAGCATAGTCAATGTCGAGCCAGCACCCAG 4135
Qy 1312 ACGGCCATAGTACAGCTGTTTGCAGGGGAACAATGCTCAGAAAGAGGTTGGCCACCTCAC 1371
Db 4136 ACAGTGATGATGAGCTGCTGGTGGGCATCAAGCAAAATCTCAGCATTTGGCCCTCGCAT 4195
Qy 1372 AGAACACTCGGTGTCTACGGAAACAAAGTCAATGTTTGTGTAATCTCTCATAGGTCTCA 1431
Db 4196 CAGATCTCTACGGTCACTCGTGAAGGGTCAATTTGCCCTGTCTGCCCCCTCGATGATTCAC 4255
Qy 1432 GCTCGGTATGATCAAGATACCTGATATCTGAGGTGTGATTTTTCAGTCAATATTTGG 1491
Db 4256 TCGAGATATGAGAAATCGCATCTACTTTCCGGCGGCATCGAATTTCCGCGGTGTGTGTG 4315

Qy 1492 GACATATTTTCTCGAGAAATGAAACATATCTTCTGTGTTTCATGTTGTGTGAGATTACTCAA 1551
Db 4316 GATCTGTACAGCGGAAGCTTTACTGTCATAGGTTTTTGGCTTTTCATGCGGTGAGATCACAA 4375
Qy 1552 CTTCCTAGTTCCACTCGAAACTGTAGTGAAGAGTACAGCACTGCACTCTCTCTGTGAGCC 1611
Db 4376 CTCTGTGTGCGCCAGAGAGCTGTAGTCCAAAGATCTCTGAAATGTCATATGTTTGGTGCC 4435
Qy 1612 AGTGACCACTCAGTAGGACTTCTTAAGTTTTCGAGAGAAAAATGTCATAATTTTGGCACT 1671
Db 4436 TCGGACCAATTCGTTACGCTAGTGTCCCTTCAGGAGCGCAAGTGTGTGACCTTGGCAAGC 4495
Qy 1672 CGTCACCTTTTCTTCTATTTCAAGTAATCAAAATGAGAGCCTTCTGATGATTTACTCTGGTGTG 1731
Db 4496 AGACATCTATTCTCTGTGTGTACCATTAATGCGCTCCCTTGGACGACTTTCTTAATCGTA 4555
Qy 1732 GGGTGTTCAGATGTTCTGTGTACGCTGCGCAATGGAATGATCTGTGTCATTTGGATCGTTGT 1791
Db 4556 GGTCTTCCGATGGCAGCGTGTATGTGTGCAATGGAACGGGTCACTCGATCGAGTG 4615
Qy 1792 GTGATGGGATAACAGCAGTTGAGATTTCTAAACGCTTTGTGATGAA----- 1836
Db 4616 CTTTCAGCGGATCTGCGCAGAGGAAGTCTTTTCGCTTGCATGAACAGGCGGAGGATGCG 4675
Qy 1837 -GCTGTTCTGCTGTGTGTGATTCACTTAGTCACTCAGCAGTCAACCTAAACAGAGTAT 1895
Db 4676 GGTCTGCGGCGCGCGGTGTGTTCCAAATGTTGTGTCATCTGCCAGCAGATGGGTATGGCCAA 4735
Qy 1896 GACGAGACGTAGTCTTCTCTCTTAAATAATGCG-----CCATCATAG 1941
Db 4736 CCGCGAGTGACCTTCTTCGCTGGCCTGGAAGTCCGCAACATGAATGCAATCCGTCATGCT 4795
Qy 1942 CTACAAACCTTGCACAACTAACCTTCTGGCTTCTGAGGCATCTGACAGGGA----- 1994
Db 4796 ACCAGCGGGGATTAGCBACTTCAGCAACTTCAGGGTCAATACCAAGGAACTTCGAC 4855
Qy 1995 --TTTACCTAAATATTTCTCATAACTCCTGATGGTTCAGCAATAAAGACAAACCTTAA 2052
Db 4856 TTTCTGATGAACAACCGCAGCAATCTTTAGTGTATCCAGGGTCTTCTGTAATAATCTTAAG 4915
Qy 2053 GACCCGACATACATCTGCTTATTTTGTGATGTTGAGCGTGTGATTTTCACTCTCTGACT 2112
Db 4916 GACCGGAGAGCCACATCTTTTCTTGTGATCGAGGACTGATCTTCGAAATTTGCATAGC 4975
Qy 2113 GAAGAAGCTCTAGGCGCAATCTGCTCTTATTTCCAGAGAAATTTGCAAAAGCATCT 2172
Db 4976 GAGGATAACGACAGATGACACAGCTACCTTGGATCACTTGGTGTCCACTGCGAAT 5035
Qy 2173 GGCAGTTTCAGACAAAGGGGCTCTTTTAACTGGAAGAACGAGCAGAGTCTTCTTCCAA 2232
Db 5036 CCAAGAGTGGCAAAATCGATGCAATTTGGATGCTAGCAAGAGATCGGTGACTTCTTCAAC 5095
Qy 2233 CAGTCAAGAGAACGATCAAGAGACATCAAGGAACACTCTTGTGATGATGAGAGGAG 2292
Db 5096 AAGTGAAGAACAGGCCGTGATGTGGAAGAGATCTCTGAAGGACAAAGCAACACGGG 5155
Qy 2293 GATGAGGAGATATGAGGACAGAGAGGGA----- 2321
Db 5156 CTGCTCAGAGTTCAAGGAGAAATCTGAGATTTCTGAGAGAAAGTACAGGCCAAGGTG 5215
Qy 2322 -----AGAAAGTGAATCTGAAATATCGGTCCAGCAAAATCAAGGCCA 2361
Db 5216 GAGAGCCTGCAAAAGGACGATAGAGCCGACGAGGAGCAGCAGGACCTGAAAGAGCAAAATC 5275
Qy 2362 TTGACCTTATGAAATATAATTTAACTATGGAACACTGCAAGCTGTTATGCTCTGCTT 2421
Db 5276 GCTTCCAAGATGGAGTTACGCAATGTTATGAGGATGAGCCAGCTACTTCTTCTGCTGCTC 5335
Qy 2422 CAGCCTGCGGTTGAAATGAACTGAGTGAAGTTTGTCTGATCGCTGATCGCTTGGAAATGCTG 2481
Db 5336 CATTTTGGGCGCTAGATCCACATCTAGACAGATGTGCGAGAGCGGCTTGGGCTTCTT 5395

XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13190.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB62133.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 13190; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 13668 BP; 3652 A; 3342 C; 3582 G; 3092 T; 0 U; 0 Other;
Query Match 8.8%; Score 393; DB 4; Length 13668;
Best Local Similarity 46.6%; Pred. No. 6.2e-107;
Matches 2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;
QY 1 ATGGCAGGAAACAGCCCTTGTCTACCCATTGTTCTTTGGGGTGGAAAGCGCCACACAT 60
DB 3491 ATGGTGAGTACCAATCTGGTGGTACCGGTGCTGTGGGGTCCACCGCGCCAGCGAC 3550
QY 61 TGCATCTCAGCGGTACTTTTAAAGATGATGGGGCCAGATCGTAAAGATGTCAGAC 120
DB 3551 TGCATCTCAGCGGTGTTCTCTCGAGCATGATTTACCTTGTCTACCGGCTCTACGAT 3610
QY 121 GGACAAATATGCTCTGGGATCT--TTACAGTAACTGCAAAATTAATCTCGAGCATG 177
DB 3611 GGACAGATCTGCTTTGGCAAGTGGAGGCCACCACTCTGAAGATGTCGGCGGGTGCCTC 3670
QY 178 TTGTTTGGTCTACAGCATCAATCACTTGTCTTAAAGCTTGTCTTCCAGTGACAA 237
DB 3671 TTGTTGGCCATTCGGCGCCCGTTTGTGCTTAGTGGTGGCTCGCTTCTGCGGAGAC 3730
QY 238 CAGTATATTGTGATGATCTGAAAGTGGAGAGATGTCCTCTGGAGTGTGATGTC 297
DB 3731 AACTTCTGCTGAGTCTGTCGAGACGGAGAGATGTGCACTCTGGATCTTACGGACGGC 3790
QY 298 AGATGATTAATTAATAAATTAAGTCTTGACACATCTGGCATACAGTTCTACAGTTC 357
DB 3791 AAGTGATGAGGGGTCAAGCTTCCACAGGTGCACACCCCAAAATTCAGAGCTACCAACC 3850

QY 358 TCTGTTGGGATCAGCGAAGAGGCTTTTATGCCAGGACATTACCTGAAATCTT 417
DB 3851 -----GCCAACAGCGAGGATGTGCGCTCTTTTGCATTTGGCTACTATGCTGAATCATG 3904
QY 418 GTTGTGATGCTACAGCCCTTGAAGTATTATATCTCTTAGTATCAAGATATACACAGAC 477
DB 3905 GTCATGATCCGTTTCAAGCTTAGAGGTGATCTAGTCTGCTGCTCAAGGTCAACCGGAT 3964
QY 478 TGGATTAGCTCCATGATATTATTTCGATCCCAAGCAACCAAGAGGACACAGTGGTAGCA 537
DB 3965 TGGATATCCGCTATTACCGTACTTGCCTCCCTCCGCAAGGATGACGTAGTGTGGCC 4024
QY 538 CTCTCGTGAATGCGCATCTGAAGTCTGATGTTTACCTCGGAAATAGTGCATGACAG 597
DB 4025 ATCACCAACCCAGGAAACCGTGAAGGTCTGAGCTCTGAC-----GGGTAAACGAGAACAA 4078
QY 598 GATACCTGAGCCAAATTTTGAAGGAGGAATCCAAACCAATTTATTGTGAGATTCGCAAGC 657
DB 4079 CACGCGAGCCCAATTTACGAGAACGAATCCAGGAGATCGCTGCTGATGCCATCACC 4138
QY 658 ATCTCTTTTGTGCTATTACACAAAGGTCACTTTTGGTTGTGTGTTTCCAAATATTGGAGG 717
DB 4139 ATGAATTTGCTGGCCAGCAACAGCGCACAGTCTCTTAGTTTGTACCAATATTGCGAG 4198
QY 718 GTGTTGATGCGGAGACTATTCTTGTGTTTCAAGTCTTAGTGAATAATGACAGACA 777
DB 4199 ATCTACGATGCGCGCACTTCACTGTACTCTGCTCGTAAATTCCTCCGCAAGGAAACGT 4258
QY 778 TGGACCGGGGGGACTTGTCTCATCAGATAAAGTATCATTTTGGACAGAAATGGCAA 837
DB 4259 TGGACGGCGGCGACTTCACTCCGACCGGGTAAATGCTATGGACAGATGAGGAAAG 4318
QY 838 AGTTATTATTACAACTACCTGCGAGTGTCTTCCAGCTAGTGAATTCATTCGCGAGTGA 897
DB 4319 GGTACTTGTACAGCTGCCGCAAACTGATTCGACACAAAGGAGTTCCATCGAAG 4378
QY 898 GTGGGGAAGGAGTGAATAATT-----TAAATTCCTCTGTACAAATATCTCTTGGATC 952
DB 4379 AGTGTGTCAGGAGTGCACCTTACCTACTACTGTGTGCAACACGCGGTGATAAGTG 4438
QY 953 GAAAGATAAAGTGTCTAATTTGCTCTCTGTTACTCGTTCCTTCTATAGTATGACAG 1012
DB 4439 CTGCTGCTGCCCCAGCCAGTGAAGTTGCTTCAAGGCGCGGTGGCCAGCAATCTACTG 4498
QY 1013 AATATTTCATAAACTGTTAAATTCAGGGTGAATCTTCTGGAAGTTTGAATATTGGAACA 1072
DB 4499 CGCGGAGACTCCGAGGAGTACATCTCTGTCTGGAATGTGCGGAGGTGCGGTAGACAT 4558
QY 1073 TATCAGACACAGCTGATTAACAGGGAAGTGAAGAAGGGCTGGCAATGACAACT-TCTATT 1131
DB 4559 AATTAGCATCTGCAAGCCCAACAGATGCGCGCGACCCCTCAAGCGCGACGCTGTGCAT 4618
QY 1132 AGTTTGCAGAGGCAATTTGATTAACCTGAATCTTGTCTGCTGGAATTAAGATCAGCTG 1191
DB 4619 TCGTTTGCAGGCGGTGTCATTAATGGAATCCACCAAGTGGGCACTCTTGGATCAGC-- 4676
QY 1192 AGTGTGATTCCCAATAGTAACTCTTAAAGTAACTGCAAGTGTGTACATACACAGCA 1251
DB 4677 -----TCTCCGCAATACAGATTCGCGGTGAGAGTCACTCTCGAGCATCTATCTGCCACAG 4732
QY 1252 CATGGAACGATTTGTTGTTGCTGTAAGATGGAAGCATAGTTATTGTGACTGCGCACACAG 1311
DB 4733 CAGAGTCCGCTGCTGATCGGACGCGAGGATGGCAGCATAGTCATAGTGGCAGCACCCAG 4792
QY 1312 ACGGCATAGTACAGCTGTTGCAAGGGAACACATGCTCAGAGAGGTTGGCCACCTCAC 1371
DB 4793 ACAGTGATGATGAGCTGCTGTTGGGCACTCAAGCAAACTTCAGCGATTGGCCCTCGCAT 4852
QY 1372 AGAACACTCCGTTGTCATCGGAACAAAGTCAATGTTTGTCTATATCTCTCATCAGGTCTCA 1431
DB 4853 CAGATCTCTACGGTCAATCGTGAAGGGTCAATTCGCTGCTCTGCCCTCGATGATTCAC 4912
QY 1432 GCTCGGTATGATCAAAAGATACCTGATATCTGGAGGTGTGGATTTTTTTCAGTCAATTTGG 1491

Db 4913 TCAGATATGAGAAATCGCATCTACTTTCCGGCGGATCGATTTCCCGGTGTGTCTGTGG 4972
QY 1492 GACATATTTCTGGAGAAATGAACAATATCTCTGTGTTCAATGGTGGTGGATTAATCTCAA 1551
Db 4973 GATCTGTACAGCGGAAGCTTACTGTCATAGTGTGTTGCTTCATCGCGGTGAGATCAACAA 5032
QY 1552 CTTCTAGTTCCACTGAAATCTTAGTGCAGAGTACAGACATGATCTGCTCTGTAGCC 1611
Db 5033 CTGCTGTGCGCGCAGAGAGCTGTAGTCCAGAAATCTGAAATGCAATATGTTCCGTGGCC 5092
QY 1612 AGTGACCACTCAGTAGGACTTCTAAGTTTGCAGAGAGAAATGCAATATGTTGGCATCT 1671
Db 5093 TCGGACCAATCCGTTAGCTAGTGTCCCTTCAGAGCGCAAGTGTGTGACCCCTGGCAAGC 5152
QY 1672 CGTCACCTTTTCTTATCAAGTAATCAATGAGGCGCTTCTGATGATTAACCTGTGGTG 1731
Db 5153 AGACATCTATTCCTGTGGTTACCAATTAATGCGTCCCTCGACGACTTTCTTAATCGTA 5212
QY 1732 GGTGTTCAGATGGTTCGTGTGACGTTCTGGCAATGATCTGGTGCATTTGATCGTTGT 1791
Db 5213 GGTGTTCGATGGCAGCGTGTATGTGTGCAAAATGGAACCGGTCACTCGATCGAGTG 5272
QY 1792 GTGATGGGATACAGCAGTTGAGATTTCTAAACGCTTGTGATGAA----- 1836
Db 5273 CTTACAGGATGCTGGCAGAGGAGTCTTCCGTTGCGATGACAGCGGAGGATGGC 5332
QY 1837 -GCTGTCTGCTGCTGTTGATTCACCTAGTCAATCCAGCAGTCAACCTAAACCAAGCTAT 1895
Db 5333 GGTGTGGGCGCGCGTGTCCCAATGGTGCATCTGCCAGCGAGATGGGTATGGCAAT 5392
QY 1896 GACGAGCTGTCTGCTGCTCTTAAATATGCG-----CCATCATAG 1941
Db 5393 CCGGCGATGCACTTCTTCGCTGCGCTGAAAGTCCGCAACATGAAATCGCATCGCTATGCT 5452
QY 1942 CTACAAACCTTGCACATCACTCTGGCTTCTGAGCCTCTGCAAGGGA----- 1994
Db 5453 ACCAGCGCGGATTAAGCACTCCAGCACTTCAGGCTCAATCCAGGAACTTCGAC 5512
QY 1995 --TTTACCTAAATATTTCTATAACTCCCTGATGGTTCAAGCAATAAAGCAAAACCTPAACA 2052
Db 5513 TTTCTGATGAACACACGCGCAATCTTTTAGTGATCCAGGCTTCTGTAATCTTAAG 5572
QY 2053 GACCGGACATACATGTGCTATTTTGTATGTGGAAGCGTTGATTAATCACTCTGACT 2112
Db 5573 GACCGGAGAGCCACATCTTTCTTTGACATCGAGGCACTGATCTTCGAAATTCGATAGC 5632
QY 2113 GAAGAAGCCTTAGGCGGAATCTGCTCTTATTTCCCGAGAGAAATTTGCAAAAGCATCT 2172
Db 5633 GAGGAATACGCAAGATGACACAGCTACCTTGGAAATCACTTGGTGTCCACCTGCGAAT 5692
QY 2173 GCGAGTTTCAGAAAGGGGCTCTTTTAACTGGAAACAGCAGCAGTCTCTTTCCAA 2232
Db 5693 CCCAAGATGGCAAAATCGATGCAATTTGGATGCTAGCAAGAAGATCGTGACTTCTTCAAC 5752
QY 2233 CAAGTGAAGAAACGATCAAGAGAACATCAAGAACACCTCTTGTATGATGAAGAGGAG 2292
Db 5753 AAGGTGAAGAACAGGCGGTAGATGTGGAGAAGATCCTGAAGGACCAAGCAAAACACGGG 5812
QY 2293 GATGAGGAGATAATGAGCAGAGAGGGA----- 2321
Db 5813 CTGGTGCAGAGTTCAAGGAGAAAATGAGATTTGTGAGAAAGATACAGGCCAAGGTG 5872
QY 2322 -----AGAAAGTATCTGAAATATCGGTCCAGCAAAATCAAGGCCA 2361
Db 5873 GAGAGCTGCAAAAGGCGAGTGGCCGACGAGGAGCAGCAGCACTTGAAGAGCAAAATC 5932
QY 2362 TTGACCTATTAGAATAATTTAACTATGCACTGCAAGCTGTTTATGTTCTGCTT 2421
Db 5933 GCTTCCAAAGTGGAGGTTACGATGTTATGGAGGTAGCCAGCTACTTCTTTCGCTGCTC 5992
QY 2422 CAGCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTGGCTCGATCGCCTTGGAAATGCTG 2481

Db 5993 CATTTCTGGGCGCTAGATCCACATCTAGACAAAGATGTGCGAGACGCGCTTGGGCGCTGCTT 6052
QY 2482 AAACCCCACTGCACCGTATCTGTTTGGCCTTGTCTCAAGAGGAGCCATATGTCACATGATG 2541
Db 6053 AGACCCATTTGTGCCATCTCATACGAGAGTCTTCCAGGCGCGCTATATGTGCTGCTG 6112
QY 2542 CTGCGGGTTATA-----TCAGCCTGCTGTGTAACCTGTCAAC 2578
Db 6113 CTGCCACATGGCAGAACTATATGCCATACACCGGTATCAAGCTTCCCTCCAGTTCC 6172
QY 2579 ATGGGAAAAACAGAGTGAAG-----GAAGCTGCCAGCGCTGAGGAGTA 2625
Db 6173 AAGAAGCGCCGCTGCCAGAGAGCTGCAGCGCTTGGAGCATCTCAAGCTGTTTTCACA 6232
QY 2626 GGAAGGGAACCTTACGAGTGTCCGTCGCTGACCAACACAGCATCTCTGTCTATCAT 2685
Db 6233 TCCGCTCTGCAATGGGAGCTCAGCACTACGTTTGAACCAACCAATATTTCTGCGCTGG 6292
QY 2686 TCTTTGGCAAAATACTTTATGATGATGACCAATGCAACTTTTATTTGGTGTATCATATAAG 2745
Db 6293 GCAATGTCANATACGCTGCTGTGATGAGTGGGCTCTCTTCTGCCGACAGCGAGAG 6352
QY 2746 AAGGCTCTTACAGGCGCACTAGACCAAGCACCCAGACCTTTTCTAAGCAAGGGGTTCC 2805
Db 6353 CACAAGAGCTGCAG-CGCTTGGCCAGCGCACGACTCCACATTCAGCAACGAGAGGA 6411
QY 2806 CCTCCAACTTCAGTAAATTTGTGACAGCAGATTAACAAAGTTGCTGCACTGCTGCTT 2865
Db 6412 GCGCGAGGAGCTGATGGCCACACATCTCTCAATCAAGCAGCGCTGAGCGCTGCTTGC 6471
QY 2866 TCGCTCGCTCTGATGCTGATCACTCTGCTCTGACCCCTCTTCTGCTCTGCTTTCAT 2925
Db 6472 TACCACCAATTCCTTCTGCTGCAAAAATGAGGCGCTGGAGCCGAGAAATTCAA 6531
QY 2926 ACCTGTTTCTTAGTAAATGAAGTTGGAGTCAAGTCTAGTCTGATGCACTGTGTATGCTG 2985
Db 6532 GGTTCGCGAGGTGGAAATGATGTCAGCGCTGGCAGCATCACTGTCATAGAG--ATCAG 6589
QY 2986 CAGACCTACTGGATGGATAAATTTAGGCTTCCCTTCTCGAGATGCTGCCCCAAGA 3045
Db 6590 GAGGCGGCTCAGCAGATCTTTTGGCGAGCTTACCCGATGGGCAAGAGGTCCGAAG 6649
QY 3046 TGGCAAGATCGATGCTTGGAGGTGAGAGAAAGCGCACAGCGCTTCTTCTGCGGAATCTG 3105
Db 6650 CAGCTAGTCGAGCTGGGCACAATATCTGCCACTCTACACGACACCGAGCCCATAGTG 6709
QY 3106 AGAAGAAATTTGACGAGCAGCAGGAGGAAGCAATGATGCTGCGCTCTTACTTACCT 3165
Db 6710 GGAGCCCAACAGCAGTTGGCTTGTAGTCAAGCGGCAAGCGGAGAGCTGGCTCGGA 6769
QY 3166 CAGTACATAGACCATCATATCACTGAGTCAATCAGAGCGCGCAGACTATCACC 3225
Db 6770 TCGGCGGCAATGGAGGGGTGGCGTCCGAGTATCCGAGGCGGTGGTCTGTTAGTGGT 6829
QY 3226 ACGGCTCTGATGCTCAGGCGCTGAAGCAAAAGTCCAGAGGAAGAGCATGACCTTGT 3285
Db 6830 AGCGGCGAGGTGGAGCGTCTCTGGGGAGATGCCCAACAGATGAAGTACAGAGAG 6889
QY 3286 GACGATGACATCACTGCTGTTTATCAAGTGTCCCAAAATGAAAAAATTTCTTAC- 3344
Db 6890 GAAGAGGAGGAGATTATACGCAAAACCATCCAGCTTGTGCGAGCTGAAGCGCAAAACT 6949
QY 3345 -----ATCTTACAGAGAAAGCAGGAGCAAGCTACCGCTATTTGTTTACTTCCAGT 3395
Db 6950 ACGGCATTTATCTTCTGGTGTATAGGAGCAGAAATTTGGCCAGGATATATCCCAGGAG 7009
QY 3396 AATAGGAGCTGAATTTG--GTGCTGAAATTTGACCTCTTAAACTATTTGACCAAGCACTCGA 3453
Db 7010 TCACCAATCAACGCTGGAAGTATTAGCATGGCCACGGAGCAATCTAACCATGGAGTT 7069
QY 3454 AGCTTACGCAAAATCTCTGAGGAGTTCCGGTTGATAGTGTGGTGGATCCCACTACTCGCTG 3513
Db 7070 GCTGGCGGGAACGTCGAAAAATCTAGTGTCTGGAGGGCTTTCGGCATAGCAACAATCTA 7129

Search completed: July 2, 2005, 04:15:25
Job time : 2170 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 00:59:56 ; Search time 18357 Seconds
(without alignments)
11806.950 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcaggaaacagcttg.....agcaccgcttcagtcttaa 4473

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4473	100.0	4473	9	AY099325 Homo sapi
2	4469.8	99.9	4548	6	AX244640 Sequence
3	4258.8	95.2	4371	9	AF188125 Homo sapi
4	3549.8	79.4	4497	6	AX244642 Sequence
5	3546.8	79.3	4467	10	AF188124
6	3537.8	79.1	4467	10	AF105813
7	3537.8	79.1	4494	6	AX244638
8	3497	78.2	6224	6	AR541843
9	3485	77.9	6072	9	AB011113
10	3382.8	75.6	4371	10	AF192379
11	3369.2	75.3	4371	10	AF188123
12	2807.4	62.8	3316	6	AX714633
13	2807.4	62.8	3316	9	AK057141
14	2680	59.9	5905	10	AK122304
15	1941.8	41.2	4460	9	BC050352
16	1788.8	40.0	3154	5	BC073055
17	1164	26.0	1436	6	CQ720875
18	969.8	21.7	1916	6	CQ719779
19	773	17.3	164076	9	AC008006

C	20	773	17.3	189371	2	AC090230	Homo sapi
C	21	613	13.7	183544	10	AC102132	Mus muscu
C	22	603.4	13.5	215898	2	AC096428	Rattus no
C	23	491	11.0	6104	9	BSM808720	Homo sapi
C	24	487.4	10.9	5900	9	BSM805990	Homo sapi
C	25	474.6	10.6	2578	6	AX833927	Sequence
C	26	474.6	10.6	2578	9	AK096055	Homo sapi
C	27	395.2	8.8	48287	2	AC017774	Drosophil
C	28	393	8.8	5221	6	CQ598374	Sequence
C	29	393	8.8	6049	3	BT015183	Sequence
C	30	393	8.8	8411	6	CQ598373	Sequence
C	31	393	8.8	13668	6	CQ578837	Sequence
C	32	393	8.8	169618	3	AC105055	Drosophil
C	33	393	8.8	170490	3	AC104147	Drosophil
C	34	393	8.8	300933	3	AE003422	Drosophil
C	35	382.8	8.6	29352	3	DMC86E4	Drosophil
C	36	293	6.6	72383	2	AC101173	Mus muscu
C	37	287.2	6.4	5292	9	BSM808902	Homo sapi
C	38	253	5.7	171185	2	AP002424	Homo sapi
C	39	253	5.7	171747	2	AC087678	Homo sapi
C	40	253	5.7	177097	2	AP001569	Homo sapi
C	41	253	5.7	179726	9	AC007052	Homo sapi
C	42	253	5.7	182411	2	AC090408	Homo sapi
C	43	226.8	5.1	211466	9	AC012301	Homo sapi
C	44	218.2	4.9	2046	3	AY051974	Drosophil
C	45	208.2	4.7	250880	10	AC102040	Mus muscu

ALIGNMENTS

RESULT 1
AY099325
LOCUS 4473 bp mRNA linear PRI 10-JUL-2003
DEFINITION Homo sapiens rabconnectin-3 beta mRNA, complete cds.
ACCESSION AY099325
VERSION AY099325.1 GI:30313408
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4473)
AUTHORS Kawabe,H., Sakisaka,T., Yasumi,M., Shingai,T., Izumi,G., Nagano,F.,
Deguchi-Tawarada,M., Takeuchi,M., Nakanishi,H. and Takai,Y.
TITLE A novel rabconnectin-3-binding protein that directly binds a
GDP/GTP exchange protein for Rab3A small G protein implicated in
Ca(2+)-dependent exocytosis of neurotransmitter
JOURNAL Gense Cells 8 (6), 537-546 (2003)
MEDLINE 22671656
PUBMED 12786944
REFERENCE 2 (bases 1 to 4473)
AUTHORS Kawabe,H., Nakanishi,H., Shingai,T. and Takai,Y.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Biochemistry and Molecular Biology, Osaka
University, Yamada-Oka 2-2, Suita, Osaka 565-0871, Japan
FEATURES
source
1..4473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..4473
/note="brain enriched protein"
/codon_start=1
/product="rabconnectin-3 beta"
/protein_id="AA033134.1"
/db_xref="GI:30313409"
/translation="MAGNSLVLPVILWGRKAPHCISAVLLTDDGATIVTGDHGGIC
LWDLVBLINPRALLFGHTASITCLSKACSSDKQYIVSASESMCLWDSGRCI
EPTKLACTHTGTFQFVGNORGLRCHGHPILVVDATSLVSLVSLKSPDW
ISSMSIIRSHRTOEDTVVALSVTGILKQVITVSEISDMQDTEPIPEESKPIYCNQCQ
SISCAFTQSRLLVCSKYNRPVFDAGVSLCSPGSENGQTWGTGDFVSSDKVILWTE
NQSQYIKLPLASCPLSDFSDGKAVENLIPPVQHILLDRKDKELLICPPVTRFFY

GCREYFKLLIQDSSRLNIWNISDTADKQSEGLAMTTSISLQEAFFDKLNPQPCAG
IIDOLSVIPNSPELVATSVYIPAHGSRVCGREDGSIIVIPATOTAIVOLLOQHEML
RWGPHRTLRGHRNKVTCLLPHQVSRAYDORYLISGVDFSVIINDLPSSEMHHIF
CVHGGBITQLLVPPENC SARVQHCISVASDSHVGLLSREKXIMLARHULFPPIQVI
KNRPSDDYLWVGSDSVYVQMDTGALDRVCMGTAVFELNACDEAVPAVDSLSPH
AVNLKQAMTRRSILAALKNMHHKLOTLATNLLASEASDKGNLPKYSHNSLMVQAIKTN
LTPDOILHVPDFVBEALIIOLLTEASRPNTALISPENLOKASGSDSGKPLTLLFVNLTMD
VLPQOVKEITKENIKELHLLDREDEEINRORRESDPEVRSKSKPLTLLFVNLTMD
TAKLFMSCLHANGLNVLDEVCLDRGLMKLPCHTQVSGLLSRGSHMSLMLPGYNQAC
KLSHGKTEVRKLPASGEVKGTYGVSRAVTTQHLHSIIISLANTLSMNTNATFIDGHH
KKGPTPRPSPDLKARGSPPTSSNI VQGOIKQVAPVVSARSDDSDPPSAP
ALHCTFLVGNESQLAAMHCVMPLDGLDKFRPPLLEMLARRWODRCLEVEAREAQV
LLAELRRIEQAQRKEALDAWAPVLOVIVDHVSPGVTSAAOTITTPADSPAERAKV
EEBHDIVDDIITGCLSSVPOMKKISTSYEERKQATAVLLGVIGAEFGAIEBPKL
LTPRPSQIPEGFGLTSGSNYSLARHTCKALTELLLOPPSPKLPHPHSIRRTALDI
LHGFTVTPPEVSJLVMLGSELCAKOLANITMGLPLSPADSAARSARHSLDI
ATVAPPAFTITAKYVHRHTALAAANTQSQNMHTTTLARAKGEILRVIBILIERKPTD
VDDLVEVMDIIMYCLEGSLVKKGLQCFPAICRFYMSVYVERNHRIAVGARHGSVA
LYDIRGKQOTLHGKGPITAVAFAPDGRYLATYSNTDSHISFWMQNTSLLGSIQMLN
SAPQLACIKTYQVPPVPAPSPGSHNALKLARLLWTSNRNVILMAHDGKEHRFMV"

ORIGIN

Query Match 100.0%; Score 4473; DB 9; Length 4473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCGAGGAAACAGCCTTGTTCTACCCATTTGTTCTTTGGGTGCGAAAGCGGCCACACAT	60
DB	1	ATGCGAGGAAACAGCCTTGTTCTACCCATTTGTTCTTTGGGTGCGAAAGCGGCCACACAT	60
QY	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCTTAAACAGATGTCACGAC	120
DB	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCTTAAACAGATGTCACGAC	120
QY	121	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTG	180
DB	121	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTG	180
QY	181	TTTGGTCATACAGATCAATCACTTGTTTGTCTTAAAGCTTGCTTCCTCCAGTGACAAACAG	240
DB	181	TTTGGTCATACAGATCAATCACTTGTTTGTCTTAAAGCTTGCTTCCTCCAGTGACAAACAG	240
QY	241	TATATTGTGAGTGCATCTGAAAGTGGAGAGATGCTGCTCTGGGATGTGGAGTGATGCGAGA	300
DB	241	TATATTGTGAGTGCATCTGAAAGTGGAGAGATGCTGCTCTGGGATGTGGAGTGATGCGAGA	300
QY	301	TGTATTGAATTTACAAATTTAGCTTGCCACATATCTGGCATAAGTTCTACCAAGTTCTCT	360
DB	301	TGTATTGAATTTACAAATTTAGCTTGCCACATATCTGGCATAAGTTCTACCAAGTTCTCT	360
QY	361	GTTGGGNAATCAGCGAGAGAGGCTTTTATGCGACGGACATTTACCTGNAATCCTTGTT	420
DB	361	GTTGGGNAATCAGCGAGAGAGGCTTTTATGCGACGGACATTTACCTGNAATCCTTGTT	420
QY	421	GTGGATGCTACAGCCTTGGAAGTATTATATCTCTTAGTATCAAGATATCACCAGACTGG	480
DB	421	GTGGATGCTACAGCCTTGGAAGTATTATATCTCTTAGTATCAAGATATCACCAGACTGG	480
QY	481	ATTAGCTCCATGAGTATTTATTCGATCCCAACGACACAGAGGACAGTGGTAGCACTC	540
DB	481	ATTAGCTCCATGAGTATTTATTCGATCCCAACGACACAGAGGACAGTGGTAGCACTC	540
QY	541	TCGGTGATCGGCATCTCGAAGGCTTGGAATTTGTTA CTTCCGAAATTAAGTGACATGCAGGAT	600
DB	541	TCGGTGATCGGCATCTCGAAGGCTTGGAATTTGTTA CTTCCGAAATTAAGTGACATGCAGGAT	600
QY	601	ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTGTGATGATTCGCAAAAGCATC	660
DB	601	ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTGTGATGATTCGCAAAAGCATC	660
QY	661	TCTTTTGTGCATTTACAAAGGTGACTTTTTGGTGTGTGTTGTTCCAAATATTGGAGGGTG	720
DB	661	TCTTTTGTGCATTTACAAAGGTGACTTTTTGGTGTGTGTTGTTCCAAATATTGGAGGGTG	720

QY	721	TTTCGATGCGGAGACTATTTCCTTGTGTGTTTCCAGTCCCTAGTGAATAATGACAGACATGG	780
DB	721	TTTCGATGCGGAGACTATTTCCTTGTGTGTTTCCAGTCCCTAGTGAATAATGACAGACATGG	780
QY	781	ACCGGGGGGAGCTTTTGTCTCATCAGATAAAGTCATCATTTTGACAGAAAATGGGCAAGT	840
DB	781	ACCGGGGGGAGCTTTTGTCTCATCAGATAAAGTCATCATTTTGACAGAAAATGGGCAAGT	840
QY	841	TATATTTTACAACTACCTGCGAGTTGCCCTTCAGCTAGTGAATTCATTCGCGAGTGTG	900
DB	841	TATATTTTACAACTACCTGCGAGTTGCCCTTCAGCTAGTGAATTCATTCGCGAGTGTG	900
QY	901	GGGAGGCGAGTTGAAAATTTTAAATCTCTCTGTAACAATATCTCTCTTGATGCAAAAGAT	960
DB	901	GGGAGGCGAGTTGAAAATTTTAAATCTCTCTGTAACAATATCTCTCTTGATGCAAAAGAT	960
QY	961	AAAGAGTTGCTAATTTGTCCTCTCTGTTACTCGGTTCTTCTATGAGTGCAGAGAAATTTTC	1020
DB	961	AAAGAGTTGCTAATTTGTCCTCTCTGTTACTCGGTTCTTCTATGAGTGCAGAGAAATTTTC	1020
QY	1021	CATAAATCTGTTAATTTACGGGTGATTTCTTCTGGAAGTTGAATTTTGGACATATCAGAC	1080
DB	1021	CATAAATCTGTTAATTTACGGGTGATTTCTTCTGGAAGTTGAATTTTGGACATATCAGAC	1080
QY	1081	ACAGCTGATAAACACAGGAGTCAAGAGGCTGGCAATGACAACTTCTATTAGTTTGCAC	1140
DB	1081	ACAGCTGATAAACACAGGAGTCAAGAGGCTGGCAATGACAACTTCTATTAGTTTGCAC	1140
QY	1141	GAGGCAATTTGATAAATCTGCTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1200
DB	1141	GAGGCAATTTGATAAATCTGCTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1200
QY	1201	CCCAATAGTAAATGAACTCTTAAAGTAACTGCAAGTGTGTATCATACCAGCAATGACCA	1260
DB	1201	CCCAATAGTAAATGAACTCTTAAAGTAACTGCAAGTGTGTATCATACCAGCAATGACCA	1260
QY	1261	CTTGTGTTGTCGTCGTAAGTGAAGCAAGTATTTTGTACCTGCCACACAGACGGCCATA	1320
DB	1261	CTTGTGTTGTCGTCGTAAGTGAAGCAAGTATTTTGTACCTGCCACACAGACGGCCATA	1320
QY	1321	GTACAGCTGTTGCAAGGGGAAACATGCTCAGAGAGGTTGGCCACCTCACAGAACATCTC	1380
DB	1321	GTACAGCTGTTGCAAGGGGAAACATGCTCAGAGAGGTTGGCCACCTCACAGAACATCTC	1380
QY	1381	CGTGTGCATCGGAAACAAAGTCAATGTTGCTATATCTCTCATCAGTCTCAGCTCGGTAT	1440
DB	1381	CGTGTGCATCGGAAACAAAGTCAATGTTGCTATATCTCTCATCAGTCTCAGCTCGGTAT	1440
QY	1441	GATCAAGATATCTGATATCTGAGGTGTGGATTTTTCAGTCATATTTTGGGACATATTT	1500
DB	1441	GATCAAGATATCTGATATCTGAGGTGTGGATTTTTCAGTCATATTTTGGGACATATTT	1500
QY	1501	TCGGAAGAAATGAAACATATCTTCTGTGTTCAATGTTGTTGAGATTA CTTCACTCTAGTT	1560
DB	1501	TCGGAAGAAATGAAACATATCTTCTGTGTTCAATGTTGTTGAGATTA CTTCACTCTAGTT	1560
QY	1561	CCACCTGAAAACCTGTAGTGCAGAGTACAGCACTGCATCTGCTGTGTAGCCAGTACCAC	1620
DB	1561	CCACCTGAAAACCTGTAGTGCAGAGTACAGCACTGCATCTGCTGTGTAGCCAGTACCAC	1620
QY	1621	TCAGTAGGACTTCTAAGTTTTCGAGAGAAAATAATGATATTTTGGCATCTCTGTCACCTT	1680
DB	1621	TCAGTAGGACTTCTAAGTTTTCGAGAGAAAATAATGATATTTTGGCATCTCTGTCACCTT	1680
QY	1681	TTTCTTATTCAAGTAAATCAAAATGGAGGCTTCTCGATGATTA CTTGCTGTTGGGTTTCA	1740
DB	1681	TTTCTTATTCAAGTAAATCAAAATGGAGGCTTCTCGATGATTA CTTGCTGTTGGGTTTCA	1740
QY	1741	GATGGTCTGTGTA CTTCTGGCAAAATGAGTACCTGTGTCATTTGGATCGTTGTCGATGGGG	1800
DB	1741	GATGGTCTGTGTA CTTCTGGCAAAATGAGTACCTGTGTCATTTGGATCGTTGTCGATGGGG	1800

1801 ATACAGCAGTTGAGATTCTAAACGCTTGATGAGCTGTTCCCTGCTGCTGATTCA 1860
1801 ATACAGCAGTTGAGATTCTAAACGCTTGATGAGCTGTTCCCTGCTGCTGATTCA 1860
1861 CTTAGTCATCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGCTCTTGCTGCTCTT 1920
1861 CTTAGTCATCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGCTCTTGCTGCTCTT 1920
1921 AAAAATATGGCCCATATAGCTACAAACCCCTTGCAACTAACTCTTGGCTCTTGAGGCA 1980
1921 AAAAATATGGCCCATATAGCTACAAACCCCTTGCAACTAACTCTTGGCTCTTGAGGCA 1980
1981 TCTGACAGGGAATTTTACCTAAATATCTCATAACTCCCTGATGGTTTCAGCAATAAG 2040
1981 TCTGACAGGGAATTTTACCTAAATATCTCATAACTCCCTGATGGTTTCAGCAATAAG 2040
2041 ACAAACTTAACAGACCCGGACATACATGCTGCTATTCTTTCATGGAAGGCTTGATTAT 2100
2041 ACAAACTTAACAGACCCGGACATACATGCTGCTATTCTTTCATGGAAGGCTTGATTAT 2100
2101 CAATCTCTGATGAAGAGCTCTAGGCCGAATACCTCTTATTTCCCTCAGAGAAATTTG 2160
2101 CAATCTCTGATGAAGAGCTCTAGGCCGAATACCTCTTATTTCCCTCAGAGAAATTTG 2160
2161 CAAAAAGCATCTGCGAGTTGAGCAAAAGGGGCTCTTTTAACTGGAACGAGCA 2220
2161 CAAAAAGCATCTGCGAGTTGAGCAAAAGGGGCTCTTTTAACTGGAACGAGCA 2220
2221 GTTCTCTTCAAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAACCTCTTGTAT 2280
2221 GTTCTCTTCAAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAACCTCTTGTAT 2280
2281 GATGAAGAGGAGTATGAGAGATAATGAGGCAGAGAAAGGGAAGGAACTGATCTGATAT 2340
2341 CGGTCCAGCAATCAAGACCATGACCCCTATTAGAAATATATTTAACTATGGAACCTGCA 2400
2341 CGGTCCAGCAATCAAGACCATGACCCCTATTAGAAATATATTTAACTATGGAACCTGCA 2400
2401 AAGCTGTTATGCTCTGCTTCAAGCTGCGGCTTGAATCAAGTACTGGAATGAATTTGC 2460
2401 AAGCTGTTATGCTCTGCTTCAAGCTGCGGCTTGAATCAAGTACTGGAATGAATTTGC 2460
2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGACCGTATCGTTGGCTCTTGTCAAGA 2520
2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGACCGTATCGTTGGCTCTTGTCAAGA 2520
2521 GGAGGCCATATGTCATGATGCTGCGGGTTAATCAGCGCTGTTGTAACCTGTCAAT 2580
2521 GGAGGCCATATGTCATGATGCTGCGGGTTAATCAGCGCTGTTGTAACCTGTCAAT 2580
2581 GGGAAACAGAGTAGAGAGAGCTGCCAGCGCTGAGGGAGTAGGAAGGGAACCTTAC 2640
2581 GGGAAACAGAGTAGAGAGAGCTGCCAGCGCTGAGGGAGTAGGAAGGGAACCTTAC 2640
2641 GGAGTGTCCGTCGCTGACCAACAGCATCTCTGCTATCATTTCTTTGGCAAAATACT 2700
2641 GGAGTGTCCGTCGCTGACCAACAGCATCTCTGCTATCATTTCTTTGGCAAAATACT 2700
2701 TTAATGAGTATGACCAATGCAATTTTATTTGGTGATCATATGAAGAGGTCCTACAGG 2760
2701 TTAATGAGTATGACCAATGCAATTTTATTTGGTGATCATATGAAGAGGTCCTACAGG 2760
2761 CCACCTAGACCAAGCAACCCAGACCTTCTAAGGCAAGGGTTCCCTCCCACTTCCAGT 2820
2761 CCACCTAGACCAAGCAACCCAGACCTTCTAAGGCAAGGGTTCCCTCCCACTTCCAGT 2820
2821 AATATTGTGCAAGCAGAGATTTAAACAAAGTTGCTGCACTCTGCTTTCCGCTCGGTCTGAT 2880
2821 AATATTGTGCAAGCAGAGATTTAAACAAAGTTGCTGCACTCTGCTTTCCGCTCGGTCTGAT 2880
2881 GCTGATCACTCTGGCTCTGACCCCTCTGCTCTGCTTTTACATACCTGTTTCTTAGTA 2940

2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTGCTTTTACATACCTGTTTCTTAGTA 2940
2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCGCAGACCTACTGCGGA 3000
2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCGCAGACCTACTGCGGA 3000
3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCCGAAGATGGCAAGATCGATGC 3060
3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCCGAAGATGGCAAGATCGATGC 3060
3061 TTGGAGGTGAGAAAGCGGCAAGGCGCTCTTCTGGCGAACTGAGAGAAATTTGAGCAG 3120
3061 TTGGAGGTGAGAAAGCGGCAAGGCGCTCTTCTGGCGAACTGAGAGAAATTTGAGCAG 3120
3121 GCAGGAGGAAGGAAGCCATTGATGCTGGGCTCTTACTTACTCTAGTACATAGACCCAC 3180
3121 GCAGGAGGAAGGAAGCCATTGATGCTGGGCTCTTACTTACTCTAGTACATAGACCCAC 3180
3181 GTCATATCATCTGGAGTCAATCAGAAAGCGCGCAGATATACCAACCGGCTCTGATGCC 3240
3181 GTCATATCATCTGGAGTCAATCAGAAAGCGCGCAGATATACCAACCGGCTCTGATGCC 3240
3241 TCAGGCGCTGAGCAAAAGTCCAGGAGGAGAGCATGACCTTCTTGAAGATGACATCACC 3300
3241 TCAGGCGCTGAGCAAAAGTCCAGGAGGAGAGCATGACCTTCTTGAAGATGACATCACC 3300
3301 ACTGGTGTCTTCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3360
3301 ACTGGTGTCTTCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3360
3361 CGGAACAAAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAAATTTGGTGTGAA 3420
3361 CGGAACAAAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAAATTTGGTGTGAA 3420
3421 ATTGAACCTCTTAACTATTGACCAACCTCTGAGAGCTCTAGCCAAATTTCTGAGGGATTC 3480
3421 ATTGAACCTCTTAACTATTGACCAACCTCTGAGAGCTCTAGCCAAATTTCTGAGGGATTC 3480
3481 GGGTTGACTAGTGGTGATGCTCAACTCTGCTGGCAGACATATTTGCAAGGCACTGACG 3540
3481 GGGTTGACTAGTGGTGATGCTCAACTCTGCTGGCAGACATATTTGCAAGGCACTGACG 3540
3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCCACACAGCACTATCCGAAGAACA 3600
3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCCACACAGCACTATCCGAAGAACA 3600
3601 GCCATTGATCTGATTTGGAGCTGGGTTTCACTGTTTGGAGGCTTACATGATGTCGCT 3660
3601 GCCATTGATCTGATTTGGAGCTGGGTTTCACTGTTTGGAGGCTTACATGATGTCGCT 3660
3661 GTTCTGATGGGCTTCTCGAACTTTGTGCGGATGCCGAGAAACAACTTGCCAAACATCACA 3720
3661 GTTCTGATGGGCTTCTCGAACTTTGTGCGGATGCCGAGAAACAACTTGCCAAACATCACA 3720
3721 ATGGGGTTGCTCTGAGCCAGCAGCTGACTGGCCCGCTCTGCGAGGCAATGCCCTCTCG 3780
3721 ATGGGGTTGCTCTGAGCCAGCAGCTGACTGGCCCGCTCTGCGAGGCAATGCCCTCTCG 3780
3781 CTCATTGCCACCGCAGACCAACCGGCTTTCATCACAACATAGCCAAAGAGGTACACAGA 3840
3781 CTCATTGCCACCGCAGACCAACCGGCTTTCATCACCACCATAGCCAAAGAGGTACACAGA 3840
3841 CATAGGCTCTTGCAGCAAAATACCAATCAGAGCAATATGCACAACTCTTGTGA 3900
3841 CATAGGCTCTTGCAGCAAAATACCAATCAGAGCAATATGCACAACTCTTGTGA 3900
3901 CGAGCTAAAGGGGAAATTTTGGAGTCAATTGAAATTTCTTATTGAAAGATGCCCAAGAT 3960
3901 CGAGCTAAAGGGGAAATTTTGGAGTCAATTGAAATTTCTTATTGAAAGATGCCCAAGAT 3960
3961 GTTGTGATCTTCTGCTGGAGGTTATGGACATCATTTATGCTCTGCTTGAAGATCTTTA 4020

Db 3961 GTTGTGGATCTTCTCGTGGAGGTTATGGACATCAATTATGTACTGCTTGAAGGATCTTTA 4020
Qy 4021 GTTAAAAAGAAAGCTCTTCAAGATGTTTCCAGCATCTGCAAGTTTCTACATGTCAGC 4080
Db 4021 GTTAAAAAGAAAGCTCTTCAAGATGTTTCCAGCATCTGCAAGTTTCTACATGTCAGC 4080
Qy 4081 TATTATGAGCGGAATCACAGAATAGCAGTTGGAGCTCGGCATGTTTCACTGGCCCTGTAC 4140
Db 4081 TATTATGAGCGGAATCACAGAATAGCAGTTGGAGCTCGGCATGTTTCACTGGCCCTGTAC 4140
Qy 4141 GACATCGGACCTGAAAATGTTCAGACATCTCATGAGACACAGGACCAATCACTGCAAGTG 4200
Db 4141 GACATCGGACCTGAAAATGTTCAGACATCTCATGAGACACAGGACCAATCACTGCAAGTG 4200
Qy 4201 GCTTTTGTCTCTGATGGAAGATATCTTGGCACCCTACTCAACACATGACGACCAATTTCT 4260
Db 4201 GCTTTTGTCTCTGATGGAAGATATCTTGGCACCCTACTCAACACATGACGACCAATTTCT 4260
Qy 4261 TTTTGGCAGATGAACAGCTCACTGCTGGAAGCATCGGCATGCTGAACCTGGCACCTCAG 4320
Db 4261 TTTTGGCAGATGAACAGCTCACTGCTGGAAGCATCGGCATGCTGAACCTGGCACCTCAG 4320
Qy 4321 CTGCGTGTCAATTAACCTACCAAGTGCCCTCTGTGAGCGCCGCTCCCGCTCCAC 4380
Db 4321 CTGCGTGTCAATTAACCTACCAAGTGCCCTCTGTGAGCGCCGCTCCCGCTCCAC 4380
Qy 4381 AATGCCCTCAAGCTGCGCGCTCATCTGAGCTTCCAAACCGCAACGCTCATCTCATGGCC 4440
Db 4381 AATGCCCTCAAGCTGCGCGCTCATCTGAGCTTCCAAACCGCAACGCTCATCTCATGGCC 4440
Qy 4441 CATGACGGGAAGGAGCAGCGCTTTCATGCTCTAA 4473
Db 4441 CATGACGGGAAGGAGCAGCGCTTTCATGCTCTAA 4473

RESULT 2

AX244640
LOCUS AX244640 4548 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 3 from Patent WO0166739.
ACCESSION AX244640
VERSION AX244640.1 GI:15859534
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Thorpeitson, S. S. and Sanders, S.
[dna encoding a tag gene (tgf-_g(b) resistance associated gene)
and its protein product
Patent: WO 0166739-A 3 13-SEP-2001;
THE DEPARTMENT OF HEALTH & HUMAN SERVICES (US)
Location/Qualifiers
1. .4548
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
76. .4548
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC88575.1"
/db_xref="GI:15859535"
/translations="MAGNSLVLPDIVWGRKAPHCISAVLLTDDGATITVGHGDIQ
LWDSVLEQNPALLFGHTASITCLSKACASDQKQIVSASESGMCLWDSVDRGI
EFTKLAHTGTIOFYQSVGNQREGRLCHGYPEILVVDATLELYSLVSKISPDW
ISSMSIIRSHRTQEDTVALSVILKLVITVSEISDMODTEPIFEESKPIYQNCQ
SISFCAFTORSLLVCSKYWRVFDAGYSLICSPSENGTWTGGDFVSSDKVIWTE
NGQSYIKLPASCLPASDSFRSDVGAENLPPVQHILLDRKXELLICPPVTFPY
GCREYFHKLIQDSSRLNINISDADKQSGEGLAMTTISLQEAFLKMLPCFAG
IIDLQSVFNSNEPLKTVASVYIPAHGRVLCRGEDGSIIVPATQTAIVOLLQGEHML
RRGWPHRTLRHNRKVTCLLYPHQVSARYDQRYLSGGVFSVIVWDFSGEMKHIF
CVHGGEITOLLVPPENCARVQHCISVADSHVGLLSLRKKCI MLASHLFPPIQVI
KWRPSDDLYLVGSGDSGVYVWQMDTGALDRVCVMGITAVEILNACDEAVPAVDSLSHP

FEATURES

source
CDS
1. .4548
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
76. .4548
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC88575.1"
/db_xref="GI:15859535"
/translations="MAGNSLVLPDIVWGRKAPHCISAVLLTDDGATITVGHGDIQ
LWDSVLEQNPALLFGHTASITCLSKACASDQKQIVSASESGMCLWDSVDRGI
EFTKLAHTGTIOFYQSVGNQREGRLCHGYPEILVVDATLELYSLVSKISPDW
ISSMSIIRSHRTQEDTVALSVILKLVITVSEISDMODTEPIFEESKPIYQNCQ
SISFCAFTORSLLVCSKYWRVFDAGYSLICSPSENGTWTGGDFVSSDKVIWTE
NGQSYIKLPASCLPASDSFRSDVGAENLPPVQHILLDRKXELLICPPVTFPY
GCREYFHKLIQDSSRLNINISDADKQSGEGLAMTTISLQEAFLKMLPCFAG
IIDLQSVFNSNEPLKTVASVYIPAHGRVLCRGEDGSIIVPATQTAIVOLLQGEHML
RRGWPHRTLRHNRKVTCLLYPHQVSARYDQRYLSGGVFSVIVWDFSGEMKHIF
CVHGGEITOLLVPPENCARVQHCISVADSHVGLLSLRKKCI MLASHLFPPIQVI
KWRPSDDLYLVGSGDSGVYVWQMDTGALDRVCVMGITAVEILNACDEAVPAVDSLSHP

ORIGIN

Query Match 99.9%; Score 4469.8; DB 6; Length 4548;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCAGGAAAACAGCGCTTCTTACCCATTTGTTTGGGGTGGAAAAGCGCCACACAT 60
Db 76 ATGGCAGGAAAACAGCGCTTCTTACCCATTTGTTTGGGGTGGAAAAGCGCCACACAT 135
Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCGTAACAGGATGTCCAGAC 120
Db 136 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCGTAACAGGATGTCCAGAC 195
Qy 121 GGACAAATATGTCTCTGGGATCTTTTCACTAGTAAGTGCACAAATTAATCTCTGAGCAGCTGTG 180
Db 196 GGACAAATATGTCTCTGGGATCTTTTCACTAGTAAGTGCACAAATTAATCTCTGAGCAGCTGTG 255
Qy 181 TTTGGTCATACAGATCAATCACTTGTGTTTCTTAAAGCTTGTGCTTCCAGTGCACAAACAG 240
Db 256 TTTGGTCATACAGATCAATCACTTGTGTTTCTTAAAGCTTGTGCTTCCAGTGCACAAACAG 315
Qy 241 TATATTGTGAGTGCATCTGAAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTATGGCAGA 300
Db 316 TATATTGTGAGTGCATCTGAAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTATGGCAGA 375
Qy 301 TGTATTGAATTTACAAATTTAGCTTTGCACACATCTGGCATACAGTTTCAACAGTTCTCT 360
Db 376 TGTATTGAATTTACAAATTTAGCTTTGCACACATCTGGCATACAGTTTCAACAGTTCTCT 435
Qy 361 GTTGGGAATCAGCGAGAGAAAGGCTTTTATGCCAGGACATACCTCGAATCTCTGTT 420
Db 436 GTTGGGAATCAGCGAGAGAAAGGCTTTTATGCCAGGACATACCTCGAATCTCTGTT 495
Qy 421 GTGGATGCTACAGCCTTGAAGTATTATCTCTTAGTATCAAGATATCACAGACTGG 480
Db 496 GTGGATGCTACAGCCTTGAAGTATTATCTCTTAGTATCAAGATATCACAGACTGG 555
Qy 481 ATTAGCTCCATGAGTATTATTCGATCCCAACGAGACACAGAGGACACAGTGGTAGCCTC 540
Db 556 ATTAGCTCCATGAGTATTATTCGATCCCAACGAGACACAGAGGACACAGTGGTAGCCTC 615
Qy 541 TCGGTGACTGGCATCTGGAAGTCTGGGATTTTACCTCGGAATTAAGTGCACATGCAAGAT 600
Db 616 TCGGTGACTGGCATCTGGAAGTCTGGGATTTTACCTCGGAATTAAGTGCACATGCAAGAT 675
Qy 601 ACTGAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTTGTGAGAAATGCCAAGCATC 660
Db 676 ACTGAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTTGTGAGAAATGCCAAGCATC 735
Qy 661 TCTTTTGTGCAATTTACACAAAGGTCACATTTTGGTGTGTTTCCAAATATTGGAGGGTG 720
Db 736 TCTTTTGTGCAATTTACACAAAGGTCACATTTTGGTGTGTTTCCAAATATTGGAGGGTG 795
Qy 721 TTGATGCGCGGAGACATTTCTTGTGTTTCACTGAGTCTCTAGTGAATATGACAGACATGG 780
Db 796 TTGATGCGCGGAGACATTTCTTGTGTTTCACTGAGTCTCTAGTGAATATGACAGACATGG 855

Qy	781	ACCGGGGGGAC	CTTTGTCTCAT	CAGATAAAGT	CATCATTTTGGACAGAAAAAT	GGGCAAAAGT	840
Db	856	ACCGGGGGGGGAC	CTTTGTCTCAT	CAGATAAAGT	CATCATTTTGGACAGAAAAAT	GGGCAAAAGT	915
Qy	841	TATATTTACAAACT	TACCTGCCAGTTG	CTTCCAGCTAGTGATT	CAATCCGCGAGTGATG	CG	900
Db	916	TATATTTACAAACT	TACCTGCCAGTTG	CTTCCAGCTAGTGATT	CAATCCGCGAGTGATG	CG	975
Qy	901	GGGAAGGCAGT	TGAAAAATTTAA	TTCCTCTGTACAA	CATATCCTCTTTGGAT	TGCAAAAGAT	960
Db	976	GGGAAGGCAGT	TGAAAAATTTAA	TTCCTCTGTACAA	CATATCCTCTTTGGAT	TGCAAAAGAT	1035
Qy	961	AAAGAGTGTCTAA	TTTGTCTCTCTGTTACT	CGTCTCTTATG	ATGATGATGATGATTTTC		1020
Db	1036	AAAGAGTGTCTAA	TTTGTCTCTCTGTTACT	CGTCTCTTATG	ATGATGATGATGATTTTC		1095
Qy	1021	CATAAACTGTTAA	TTTCAGGGTGATTCT	TTCTGGAAGTTG	TAATATTTTGGAA	CATATCAGAC	1080
Db	1096	CATAAACTGTTAA	TTTCAGGGTGATTCT	TTCTGGAAGTTG	TAATATTTTGGAA	CATATCAGAC	1155
Qy	1081	ACAGCTGATAAAC	AGGAAGTGAA	GAAGGGCTGGCAAT	TGACAACTTCTAT	TATAGTTTGGCAA	1140
Db	1156	ACAGCTGATAAAC	AGGAAGTGAA	GAAGGGCTGGCAAT	TGACAACTTCTAT	TATAGTTTGGCAA	1215
Qy	1141	GAGGCATTTGAT	TAAACTGATTCCT	TGCTGCTGGAAT	TATAGATCAGCT	TGAGTGTGATT	1200
Db	1216	GAGGCATTTGAT	TAAACTGATTCCT	TGCTGCTGGAAT	TATAGATCAGCT	TGAGTGTGATT	1275
Qy	1201	CCCAATAGTAA	TAAACCTCTTAA	AGTAACTGCAAGT	CTGTACATACCAG	CACATGAGCA	1260
Db	1276	CCCAATAGTAA	TAAACCTCTTAA	AGTAACTGCAAGT	CTGTACATACCAG	CACATGAGCA	1335
Qy	1261	CTTGTTTGTGCT	CGTGAAGATGGA	AGCATAGTTATT	TGTACTGCTGCCA	CA	1320
Db	1336	CTTGTTTGTGCT	CGTGAAGATGGA	AGCATAGTTATT	TGTACTGCTGCCA	CA	1395
Qy	1321	GTACAGCTGTTG	CAAGGGGAA	CATGCTCAGAA	GAGTTGGCCACT	CA	1380
Db	1396	GTACAGCTGTTG	CAAGGGGAA	CATGCTCAGAA	GAGTTGGCCACT	CA	1455
Qy	1381	CGTGGTCA	TGGAACAGT	TCACATGTTTCTAT	TATCCTCATCAGGT	CTCAGCTCGGTAT	1440
Db	1456	CGTGGTCA	TGGAACAGT	TCACATGTTTCTAT	TATCCTCATCAGGT	CTCAGCTCGGTAT	1515
Qy	1441	GATCAAAAGATA	CCTGTATCTG	GAGGTGTGGAATTTT	CAGTCAATAATTT	TGGGACATATTT	1500
Db	1516	GATCAAAAGATA	CCTGTATCTG	GAGGTGTGGAATTTT	CAGTCAATAATTT	TGGGACATATTT	1575
Qy	1501	TTCTGGAGAAAT	TGAAACATATCT	CTGTGTTTCA	TGGTGGAGATTA	CTCAACTTCTAGTT	1560
Db	1576	TTCTGGAGAAAT	TGAAACATATCT	CTGTGTTTCA	TGGTGGAGATTA	CTCAACTTCTAGTT	1635
Qy	1561	CCACTTGAAACT	GTAGTGCAGAGT	ACAGCATG	CACTGCTGTGAGCCAGT	GACAC	1620
Db	1636	CCACTTGAAACT	GTAGTGCAGAGT	ACAGCATG	CACTGCTGTGAGCCAGT	GACAC	1695
Qy	1621	TCAGTAGACCT	TCTAAAGTTT	TCGAGAGAAAAAT	TGCAATAATTT	TGGCATCTCTCGTCACTT	1680
Db	1696	TCAGTAGACCT	TCTAAAGTTT	TCGAGAGAAAAAT	TGCAATAATTT	TGGCATCTCTCGTCACTT	1755
Qy	1681	TTTTCTAT	TCAAGTAATCAAA	TGAGGCCCTTCT	GATGATTA	CTCTGTTGGTGGGGTGTCA	1740
Db	1756	TTTTCTAT	TCAAGTAATCAAA	TGAGGCCCTTCT	GATGATTA	CTCTGTTGGTGGGGTGTCA	1815
Qy	1741	GATGGTCT	GTGTAAGTCTGG	CAATGATATCTGGT	GCANTTGGATCGT	TGTGTGATGGGG	1800
Db	1816	GATGGTCT	GTGTAAGTCTGG	CAATGATATCTGGT	GCANTTGGATCGT	TGTGTGATGGGG	1875
Qy	1801	ATAACAGCAGT	TTGAGATTCT	TAAACGCTTGT	TGATGAAGCTGT	TCTCTGCTGTTGATTA	1860
Db	1876	ATAACAGCAGT	TTGAGATTCT	TAAACGCTTGT	TGATGAAGCTGT	TCTCTGCTGTTGATTA	1935
Qy	1861	CTTAGT	CATCCAGCAGT	CAACCTAAAA	CAAGCTAT	TGACGAGAGCTAGTCTTGCTGCTT	1920

Db 3016 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTTTATGCTGCAGACCTACTGGGA 3075
QY
3001 TTGGATAAATTTAGGCTCCCTTCTGGAGATCTGGCCGGAAGATGGCAAGATCCATGC 3060
Db
3076 TTGGATAAATTTAGGCTCCCTTCTGGAGATCTGGCCGGAAGATGGCAAGATCCATGC 3135
QY
3061 TTGGAGGTGAGAGAGCCGACAGAGCCCTCTTCTGGCGAAGATGAGAGAGATTTAGCAG 3120
Db
3136 TTGGAGGTGAGAGAGCCGACAGAGCCCTCTTCTGGCGAAGATGAGAGAGATTTAGCAG 3195
QY
3121 GCAGGCGAGGAGGAGCAATGTATGCTGGCTCTTCTTACTTACTCTAGTACATAGACAC 3180
Db
3196 GCAGGCGAGGAGGAGCAATGTATGCTGGCTCTTCTTACTTACTCTAGTACATAGACAC 3255
QY
3181 GTCATATCACTGGAGTCACATCAGAAGCCGCGACAGATCTACACCGGCTCTGATGCC 3240
Db
3256 GTCATATCACTGGAGTCACATCAGAAGCCGCGACAGATCTACACCGGCTCTGATGCC 3315
QY
3241 TCAGGGCTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTGTGACGATGACATCACC 3300
Db
3316 TCAGGGCTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTGTGACGATGACATCACC 3375
QY
3301 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3360
Db
3376 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3435
QY
3361 CGAAGCAAGCTACCGTATTTGTTTACTTGGAGTATAGGAGCTGAATTTGGTGTGAA 3420
Db
3436 CGAAGCAAGCTACCGTATTTGTTTACTTGGAGTATAGGAGCTGAATTTGGTGTGAA 3495
QY
3421 ATTGAACCTCTAAATCTATTGACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGATTC 3480
Db
3496 ATTGAACCTCTAAATCTATTGACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGATTC 3555
QY
3481 GGGTTGACTAGTGGTGGATCAACTACTCGCTGGCCAGACATCTTGCAAGGCACTGACG 3540
Db
3556 GGGTTGACTAGTGGTGGATCAACTACTCGCTGGCCAGACATCTTGCAAGGCACTGACG 3615
QY
3541 TTTCTTCTGTACAGCTTCAAGCCCAACTTCTTCCACACAGCATATCCGAGAAACA 3600
Db
3616 TTTCTTCTGTACAGCTTCAAGCCCAACTTCTTCCACACAGCATATCCGAGAAACA 3675
QY
3601 GCATATGATCTGATTTGAGCTGGTTCACCTGTTTGGAGCTTACATGATGTGCTCGCT 3660
Db
3676 GCATATGATCTGATTTGAGCTGGTTCACCTGTTTGGAGCTTACATGATGTGCTCGCT 3735
QY
3661 GTTCTGATGGGCTTCTCGAACTTTGTGCGATGCGGAGAAACAATTCGCAACATCACA 3720
Db
3736 GTTCTGATGGGCTTCTCGAACTTTGTGCGATGCGGAGAAACAATTCGCAACATCACA 3795
QY
3721 ATGGGGTGGCTCTGAGCCAGCAGCTGACTCGGCCGCTCTGGAGGCATGCCCTCTCG 3780
Db
3796 ATGGGGTGGCTCTGAGCCAGCAGCTGACTCGGCCGCTCTGGAGGCATGCCCTCTCG 3855
QY
3781 CTCAATGCCACCCGACAGACCCGCTTCTATCACCACCATAGCCAAAGAGTACACAGA 3840
Db
3856 CTCAATGCCACCCGACAGACCCGCTTCTATCACCACCATAGCCAAAGAGTACACAGA 3915
QY
3841 CATACGCTCTTGAGCAAAATACCAATCAACAGCAATATGACACAACTCTTGCA 3900
Db
3916 CATAGGCTCTTGAGCAAAATACCAATCAACAGCAATATGACACAACTCTTGCA 3975
QY
3901 CGAGCTAAAGGGGAAATTTGAGAGTCATTGAATTTCTTATGAAAAGATGCCACAGAT 3960
Db
3976 CGAGCTAAAGGGGAAATTTGAGAGTCATTGAATTTCTTATGAAAAGATGCCACAGAT 4035
QY
3961 GTTGTGATCTTCTGTTGAGGTTATGAGCATCAATATGACTGCTTGAAGGATCTTTA 4020
Db
4036 GTTGTGATCTTCTGTTGAGGTTATGAGCATCAATATGACTGCTTGAAGGATCTTTA 4095
QY
4021 GTTAAAAAGAAAGTCTTCAAGATGTTTCCACGCCATCTGCAAGGTTCTATAGGTGAGC 4080
Db
4096 GTTAAAAAGAAAGTCTTCAAGATGTTTCCACGCCATCTGCAAGGTTCTATAGGTGAGC 4155

QY 4081 TATTATGAGCGAATCACAGAATAGCAGTTGGAGCTCGCATTGGTTCACTGGCCCTGTAC 4140
Db 4156 TATTATGAGCGAATCACAGAATAGCAGTTGGAGCTCGCATTGGTTCACTGGCCCTGTAC 4215
QY 4141 GACATCCGGACTGGAAAAATGTGAGCAAAATCCATGGACACAAGGGACCAATCACTGCAGTG 4200
Db 4216 GACATCCGGACTGGAAAAATGTGAGCAAAATCCATGGACACAAGGGACCAATCACTGCAGTG 4275
QY 4201 GCTTTTGTCTCTGATGGAAGATATCTTGGCCACTACTGAAACACTGACAGCCACATTTCT 4260
Db 4276 GCTTTTGTCTCTGATGGAAGATATCTTGGCCACTACTGAAACACTGACAGCCACATTTCT 4335
QY 4261 TTTTGGCAGATCAACACGTCACCTGTCTGGGAAGCATCGGCATCTGAACTGCGCACCTCAG 4320
Db 4336 TTTTGGCAGATCAACACGTCACCTGTCTGGGAAGCATCGGCATCTGAACTGCGCACCTCAG 4395
QY 4321 CTGCGCTGATTAATAACCTTACCAGGTGCCCCCTGTGCGAGCCCGCTGCCCGGCTCCAC 4380
Db 4396 CTGCGCTGATTAATAACCTTACCAGGTGCCCCCTGTGCGAGCCCGCTGCCCGGCTCCAC 4455
QY 4381 AATGCCCTCAAGCTGCGCGGCTCATCTGGACTTCCAAACCGCAACGTCTCATCTGCGC 4440
Db 4456 AATGCCCTCAAGCTGCGCGGCTCATCTGGACTTCCAAACCGCAACGTCTCATCTGCGC 4515
QY 4441 CATGACGGGAAGGAGCACCGCTTCATGGTCTAA 4473
Db 4516 CATGACGGGAAGGAGCACCGCTTCATGGTCTAA 4548

RESULT 3

AF188125 4371 bp mRNA linear PRI 15-SEP-2001
LOCUS Homo sapiens TGF-beta resistance-associated protein TRAG (TRAG)
DEFINITION mRNA, partial cds.
ACCESSION AF188125
VERSION AF188125.1 GI:15624074
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE 1 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE TRAG: a novel gene associated with TGF-beta resistance
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1999) LEC, NIH, 37 Convent Drive, Bldg. 37, Rm 3C28, Bethesda, MD 20892-4255, USA
FEATURES
source
1. .4371
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21.1"
/tissue_type="liver"
1. .>4371
/gene="TRAG"
1. .>4371
/gene="TRAG"
/codon_start=1
/product="TGF-beta resistance-associated protein TRAG"
/protein_id="AA03983.1"
/translation="MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDGIQ
LWDLSEVQNPRLLPFGHTASITCLSKACASSDKQYIVSASESMCLWVDSDGRCI
ETFKLACTHTGTFQFVSQVGNOREGRLLCHGHYPVLLVVDATSLVLYSLVSKISPDW
ISMSLIRSHRTQEDTVVALSKVWITSEISDMQDTEPFEEESKPIYQNCQ
STSPCAFTQVSLVVCSSKYRVDFDGLSCGSPSENGQTWTGDFVSSDKVIWTE
NQSQSYTKLPASCLPASDSFRSDVGKAVENLIPPVQHILLDRKDELLICPPVTREFFY

GCREYFHKLII QGDSGRLNIWNI SOTADQKGSBGLANTTTSISLQEFAPKLCNCPGAP
I IQDLSVPSNPEKLVATSVIYIPHGRVUDRGDESVIYIPATQIYIQLLQGEHMLI
KRGWPHRTLRHKNRNTCLLYIPHQVYADRLQSGVDFSVIWDIFSGEKMHTI
CVCHEGIIYLLVPPECSARVQHCISVADHSVGLLSREKCIEMASRLHPFIVQI
KWRGSDIYLVQGSQSVYQMDTGRVCVMTIYIYINACDEAPVAPVLSIPI
AVNLKQATERSLAAKLIWNAHKLITQATNLLASASQKMLPKYSHNSLMQVATKTN
LTPDQIHLVDFVEALIYIQLLTERSPNTALI SPENIQAAGSGDKSGYTCGRKAA
VLFQOQVETKENIKHEHLIDDEBDEEIMQRORHESDEPYRASKPILMLPYNWD
TAKUMFSLKAWGLNEVLDVBELDZGLMKLPCYVFLSGRGGHMSPLMLPYNQAPAC
KLHGKTEVRGKLPASEGVBGKGTGYVSRAVTPQTHLLSIYSLANTLWMTNATPILGDH
KKGPTRPRSTPDLKARGSPPTSNISVQGOIKQMGSLQAMHCWMLDPLGLGFR
PPLLEMLARWQRCIYEVRAQAALLAELRIEQGRKEALDAWAPYLPQYIDHVIS
PGVTSAAQATITPADAGFEAKQBEHEDVDDITTCGLSVQPMKSTSYEERH
KQATIALGVYGAEGAEITPPLKTLRPSQSPQEGFGLTSGSNISLARHTCKAL
TELLQPPSPKPHSTIRKTAIIDLIGRGTFTWYBPMQVAVLMLGLELCAADAKQLHA
NITWGLPSPAADASAPASNAUSLITAPAPPAFTIYAEVHRHTLALANTQSQNH
TTTLARAGBILSVIELITKOMPDVVDLLVEMDOIYMCGLSVTKVQCGEFCFPAI
CFRTWYVSEYERNRIYAVGARHGSVALYDITRTCKQOTIHKHGKPTIYAVAFAPDGLZ
YSNTDSNIIYFWAHDQKSGHIGMLNASPOLRCIKTIVYPPQVQAPGSHNALKLARLI
YNTSDNRIYLMQNTDKEHRFWY"

1801 ATAAACAGAGTTGAGATCTTAAACGCTGTGTGATGAAGCTGTTCCTGCTGCTGTGATTCA 1860
1801 ATAAACAGAGTTGAGATCTTAAACGCTGTGTGATGAAGCTGTTCCTGCTGCTGTGATTCA 1860
1861 CTTAGTCATCCAGCAGTCAACCTAAACAAAGCTATGACGAGACGTAAGTCTTCTGCTGCTTT 1920
1861 CTTAGTCATCCAGCAGTCAACCTAAACAAAGCTATGACGAGACGTAAGTCTTCTGCTGCTTT 1920
1921 AAAAATATGGCCCATCATTAAGCTACAAACCTTGCNACTAACTCTTGGCTTCTGAGGCA 1980
1921 AAAAATATGGCCCATCATTAAGCTACAAACCTTGCNACTAACTCTTGGCTTCTGAGGCA 1980
1981 TCTGACAAAGGAAATTTACTTAAATTTCTCAAACTCCCTGATGGTTCAAGCAATAAG 2040
1981 TCTGACAAAGGAAATTTACTTAAATTTCTCAAACTCCCTGATGGTTCAAGCAATAAG 2040
2041 ACAAACTTAACAGACCCGACATACATATGTGCTATTCTTTGATGTGGAAAGCTTGAATTAT 2100
2041 ACAAACTTAACAGACCCGACATACATATGTGCTATTCTTTGATGTGGAAAGCTTGAATTAT 2100
2101 CAACTCCTGACTGAAAGGCTCTAGGCGGAATACCTGCTCTTATTTCCCGACAGAAATTG 2160
2101 CAACTCCTGACTGAAAGGCTCTAGGCGGAATACCTGCTCTTATTTCCCGACAGAAATTG 2160
2161 CAAAGAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTAACTGGAAACGAGCAGCA 2220
2161 CAAAGAGCATCTGGCAGTTTCAAGCAAAAGGGGCTCTTTTAACTGGAAACGAGCAGCA 2220
2221 GTTCTCTTCCAAAGTGAAGAAACGATCAAAGAGAAACATCAAGGAAACACCTCTCTTGAT 2280
2221 GTTCTCTTCCAAAGTGAAGAAACGATCAAAGAGAAACATCAAGGAAACACCTCTCTTGAT 2280
2281 GATGAAGAGAGATGAGAGATTAATGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2340
2281 GATGAAGAGAGATGAGAGATTAATGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2340
2341 CGGTCCAGCAATCAAAAGCATTTGACCTTATAGAAATAATTTAACTATGACACTGCA 2400
2341 CGGTCCAGCAATCAAAAGCATTTGACCTTATAGAAATAATTTAACTATGACACTGCA 2400
2401 AAGCTGTTTATGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 2460
2401 AAGCTGTTTATGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 2460
2461 CTGGATCCGCTTGGAGTGAACCCCACTGCAAGTATCGTTGGCCCTTGTCAAGA 2520
2461 CTGGATCCGCTTGGAGTGAACCCCACTGCAAGTATCGTTGGCCCTTGTCAAGA 2520
2521 GGAGGCGATATGTCATGCTGCTGCGGGTTAATCAGCCTGCTGTGTAATCTGTACAT 2580
2521 GGAGGCGATATGTCATGCTGCTGCGGGTTAATCAGCCTGCTGTGTAATCTGTACAT 2580
2581 GGGAAACAGAGATGAGAGAGAGCTGCAAGCTTGAAGGAGTGAAGTGAAGTGAAGTGAAG 2640
2581 GGGAAACAGAGATGAGAGAGAGCTGCAAGCTTGAAGGAGTGAAGTGAAGTGAAGTGAAG 2640
2641 GGAGTCTCCGCTGCGCTCAACACAGCATCTCTGCTCTATCATTTCTTTGCGAATACT 2700
2641 GGAGTCTCCGCTGCGCTCAACACAGCATCTCTGCTCTATCATTTCTTTGCGAATACT 2700
2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGATCATATGAAGAGGGTCTTACGAG 2760
2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGATCATATGAAGAGGGTCTTACGAG 2760
2761 CCACTTAGACCAAGCACCAGACCTTTCTAAGGCAAGGGGTTCCCTCAACTTCCAGT 2820
2761 CCACTTAGACCAAGCACCAGACCTTTCTAAGGCAAGGGGTTCCCTCAACTTCCAGT 2820
2821 AATATTTGTCAGGACAGATTAACAAAGTTGTGACCTGTCGTTTCCGCTCGGCTGAT 2880
2821 AATATTTGTCAGGACAGATTAACAAAGTTGTGACCTGTCGTTTCCGCTCGGCTGAT 2880
2821 AATATTTGTCAGGACAGATTAACAAAGTTGTGACCTGTCGTTTCCGCTCGGCTGAT 2880
2881 GCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTCTGCTTTTACATACCTGTTCTTAGTA 2940

2846 ----- 2845
2941 AATGAGGTGGAGTCACTTACTGCTATGCACTGTGTATGCTGCCAGACCTACTGGGA 3000
2846 ----- AAGGTTGGAGTCACTTACTGCTATGCACTGTGTATGCTGCCAGACCTACTGGGA 2901
3001 TTGGATAAATTTAGGCTCTCCCTTCTGAGAGATGCTGGCCGCAAGATGCAAGATCGATGC 3060
2902 TTGGATAAATTTAGGCTCTCCCTTCTGAGAGATGCTGGCCGCAAGATGCAAGATCGATGC 2961
3061 TTGGAGGTGAGAGAGAGCCGCAAGCCCTGCTTCTTGGCGGAACCTGAGAGAAATTTGAGCAG 3120
2962 TTGGAGGTGAGAGAGAGCCGCAAGCCCTGCTTCTTGGCGGAACCTGAGAGAAATTTGAGCAG 3021
3121 GCAGCGAGAGAGAGAGCCATTTGATGCTGCGCTCTTACTTACCTTACCTCAGTACATAGCCAC 3180
3022 GCAGCGAGAGAGAGAGCCATTTGATGCTGCGCTCTTACTTACCTCAGTACATAGCCAC 3081
3181 GTCATATCACTGGAGTCACTACAGAGCCGCGCAGACTATCAACACGCTCTCTGATGCC 3240
3082 GTCATATCACTGGAGTCACTACAGAGCCGCGCAGACTATCAACACGCTCTCTGATGCC 3141
3241 TCAGGCGCTGAAGCAAAAAGTCCAGGAGAGAGCATGACCTTGTGTGAAGATGATCAATCACC 3300
3142 TCAGGCGCTGAAGCAAAAAGTCCAGGAGAGAGCATGACCTTGTGTGAAGATGATCAATCACC 3201
3301 ACTGCTGCTTATCAAGTGTCCCAAAATGAAGAAATTTCTACATCTTACGAGGAAAGA 3360
3202 ACTGCTGCTTATCAAGTGTCCCAAAATGAAGAAATTTCTACATCTTACGAGGAAAGA 3261
3361 CGGAGAGAGCTTACCGCTATTGTTTTTCTTGGAGTAATAGAGCTGAAATTTGGTCTGAA 3420
3262 CGGAGAGAGCTTACCGCTATTGTTTTTCTTGGAGTAATAGAGCTGAAATTTGGTCTGAA 3321
3421 ATTGAACCTCTTAAACTATTTGACAGACCTTGAAGCTCTAGGCCAAATTTCTGAGGGAATTC 3480
3322 ATTGAACCTCTTAAACTATTTGACAGACCTTGAAGCTCTAGGCCAAATTTCTGAGGGAATTC 3381
3481 GGGTGAAGTGTGAGTGAAGTCAAGTCTTACAGAGACCTTGAAGCTCTAGGCCAAATTTCTGAGGGAATTC 3540
3382 GGGTGAAGTGTGAGTGAAGTCAAGTCTTACAGAGACCTTGAAGCTCTAGGCCAAATTTCTGAGGGAATTC 3441
3541 TTTCTTCTGCTACAGCTTCAAGCCCAAACTTCTCCACACAGACATCTATCCGAGAAACA 3600
3442 TTTCTTCTGCTACAGCTTCAAGCCCAAACTTCTCCACACAGACATCTATCCGAGAAACA 3501
3601 GCCATGATCTGAATTTGAGAGCTGAGTCTGAGTGGAGCTTACATGAGATGTGCTCGCT 3660
3502 GCCATGATCTGAATTTGAGAGCTGAGTCTGAGTGGAGCTTACATGAGATGTGCTCGCT 3561
3661 GTTCTGATGGGCTTCTCGAATTTGCGGATGCGGAGAAACAACTTGCACACATCAACA 3720
3562 GTTCTGATGGGCTTCTCGAATTTGCGGATGCGGAGAAACAACTTGCACACATCAACA 3621
3721 ATGCGGTTGCTCTGAGCCGAGCTGACTCGGCGCTCTGCGAGGATGATCCCTCTCG 3780
3622 ATGCGGTTGCTCTGAGCCGAGCTGACTCGGCGCTCTGCGAGGATGATCCCTCTCG 3681
3781 CTCAATGCGCACCGCAGACCAACCGCTTCACTACCAACATAGCCAAAGAGGTACACAGA 3840
3682 CTCAATGCGCACCGCAGACCAACCGCTTCACTACCAACATAGCCAAAGAGGTACACAGA 3741
3841 CATAGGCTCTGAGCAGCAATACCAATCAGAGAGATATGCAACACAACTCTTTGCA 3900
3742 CATAGGCTCTGAGCAGCAATACCAATCAGAGAGATATGCAACACAACTCTTTGCA 3801
3901 CGAGCTAAAGGGCAATTTTGAAGTCAATTGAAATTTCTTATTTGAAAGATGCCCCAGAT 3960
3802 CGAGCTAAAGGGCAATTTTGAAGTCAATTGAAATTTCTTATTTGAAAGATGCCCCAGAT 3861
3961 GTTGTGAGTCTTCTCGTGGAGGTATGAGCATCAATATGTAATGCTGCTTGAAGATCTTTA 4020

QY	781	ACCGGGGGACCTTTGTCTCATCAGATAAGTTCATTTGGACAGAAAATGGCAAGT	840	Db	1888	CTTAGTCATCAGCAGTCAACCTGAAACAAGCCATGACAAGACGGAGTCTCGCTGCCCTT	1947
Db	808	ACTGGAGGGACCTTTGTGTCTGACAGACAAGTCATCATCTGACAGAAAATGGCAGAGT	867	QY	1921	AAAAATATGGCCCATCAAGCTCAAAACCCCTTGCAACTTAACCTCTTGCTTCTGAGGCA	1980
QY	841	TATATTTACAACTACCTGCCAGTTGCCCTTCCAGCTAGTGAATTCATTCGCGAGTGATGG	900	Db	1948	AAAAATATGGCCCATCAAGCTGCAAAACCCCTTGCAAGCAACCTTTGGCTTCTGAGGCC	2007
Db	868	TACATCTACAACTCCCTGCCAGTTGCCCTTCCAGCTAGTGAATTCATTCGCGAGCGAGTG	927	QY	1981	TCTGCAAGGGAAATTTTACCTAAATATTTCTCATTAACCTCCCTGATGGTTCAAGCAATAAAG	2040
QY	901	GGGAAGCGAGTTGAAATTTAATTCCTCCTGTACAAACATATCCTCTTGATCGAAAGAT	960	Db	2008	TCTGCAAGGGTAATTTTACCTTAATATTTCCCAATACTCCCTGATGGTTCAAGCAATAAAG	2067
Db	928	GGGAAGCGAGTTGAAATTTGATTCCTCCTGTGCGAGCATAGCTCTTGATCAGAAGAT	987	QY	2041	ACAAACCTTAAACAGACCCGAGACATACATGTGCTATTTTGTGATGTGGAAGCGTTGATTT	2100
QY	961	AAAGAGTTGCTAATTTGCTCCTCTGTACTCTGCTTCTCTATGAGATGACAGAAATTTTC	1020	Db	2068	ACAAACCTTAAACAGACCCGAGATATCCTGTGCTTTTCTTGATGTGGAAGCTTTGATTT	2127
Db	988	AAAGAGTTGCTAATTTGCTCCTCTGTACTCTGCTTCTCTATGAGATGACAGAAATTTTC	1047	QY	2101	CAACTCTCTGATGAAAGAGCCCTTAGGGCCGAACTACTGCACTTATTTTCCCGAGAGATCTG	2160
QY	1021	CATAAATCTGTAAATTCAGGGGTGATTCCTCTGGAAGGTTGAATATTTTGGAAACATATCAGAC	1080	Db	2128	CAACTCTCTGATGAAAGAGCCCTTAGGGCCGAACTACTGCACTTATTTTCCCGAGAGATCTG	2187
Db	1048	CATAAGCTACTAATTCAGGGCGATTCCTCTGGAAGGTTAAATATCTGGAACATAGCAGAC	1107	QY	2161	CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTAACTTGAAGAAACGAGCAGCA	2220
QY	1081	ACAGCTGATAAACAGGGAGTGAAGAGGGCTGGCAATGACAACTTCTATTAGTTTGCMA	1140	Db	2188	CAGAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTCTTAACCTTGGGAGCGAGCGG	2247
Db	1108	ATAGCAGAGAAAACAGGAAGCCGATGAAGGGCTAAAGATGACAACTTGTATTAGCTTGCMA	1167	QY	2221	GTTCCTCTTCCCAACAGTGAAGAAACGATCAAAGAGAAATCAACAGGAACACCTCTCTGAT	2280
QY	1141	GAGGCATTTGATAAAGCTGAATCCTCTGCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1200	Db	2248	GTTCCTCTTCCCAACAGTGAAGAAACGATCAAAGAGAAATCAACAGGAACACCTCTCTGAT	2307
Db	1168	GAGGCATTTGATAAAGCTGAATCCTCTGCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1227	QY	2281	GATGAAGAGGAGGATGAGGAGATAATGAGGCGAGAGAGGAGAGGAGATGATCTCTGAATAT	2340
QY	1201	CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGCACATGGACGA	1260	Db	2308	CAGGAGGAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2367
Db	1228	CCCAACAGCAATGAACCTCTTAAAGTAACTGCGAGTGTCTACATACAGCACACAGGGCGA	1287	QY	2341	CGGTCAGCAATCAAAAGCCATTGACCTTATTAGAAATATAATTTAACTATGACACTGCA	2400
QY	1261	CTTGTTTGTGGTGTGAAGATGAAGCATAGTTATTTGTACTGCGCACAGAGCGGCATA	1320	Db	2368	CGGGCCAGCAAGTCCAGCCACTGACCTACTAGAAATACNACCTTACTATGATACCGCA	2427
Db	1288	CTTGTTTGTGGCGGAGATGAAGCATCAATTTTGTGCTGCGCACCCAGCAGCGGCATA	1347	QY	2401	AAAGCTGTTTATGCTCTGCTTCCACGCTGGGGTTTGAATGAAGTACTGGATGAAGTTGC	2460
QY	1321	GTACAGCTGTGCAAGGGAAACACATGCTCAGAGAGGTTGGCACCTCACAGAACACTC	1380	Db	2428	AAATATTATTCATGCTCTGCTCCACGCTGGGGCTTGAATGAAGTCTTGAATGAAGTTGT	2487
Db	1348	GTTTCACTATTTGCAGAGAAACACATGCTCAGAGAGGTTGGCCCTCACAGAACCTTC	1407	QY	2461	CTGGATCGGCTTGGAAATGCTGAAACCCCACTGACCGTATCGTTGGCTCTTGTCAAGA	2520
QY	1381	CGTGGTCATCGGAACAAAGTCAATGTTTGTATATPCTCATCAGGTCTCAGCTCGGTAT	1440	Db	2488	CTCGATCGCTCGGAATGCTGAAACCGCACCTGACGGTGTCTTGTGCTCTCTATCGAGA	2547
Db	1408	CGTGGCCATCGGAACAAAGTCACTGCTGCTGTGTAATCCCATCAGGTCTCAGCTCGGTAT	1467	QY	2521	GGAGCCCATATCTACTGATGCTCGGGTTATAATCAGCCTGCTGTGTAACCTGTCACAT	2580
QY	1441	GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCATATATTTTGGACATATTT	1500	Db	2548	GGAGTCAATGTCCTGATGCTTCCGGTTATATCAGGCTGCTGGAATACTGTGCTCAT	2607
Db	1468	GACCAAGATACCTGATATCTGAGGTGTGGATTTTTCGCTCATATATTTTGGGACATATTT	1527	QY	2581	GGGAAAACAGAAAGTAGGAAGCTGCCAGCTGTGAGGGAGTAGGAAGGGAACCTTAC	2640
QY	1501	TCTGGAGAAATGAACATATCTTCTGTTTCATGGTGTGAGATTACTCAACTTCTAGTT	1560	Db	2608	GGGAAAACAGAAAGTAGGAAGGAGCTGCCAGCAGCGAGGGGCTAGGCAAGGGAACCTAC	2667
Db	1528	TCTGGAGAAATGAACATATCTTCTGTTTCATGGTGTGAGATTACTCAACTTCTAGTT	1587	QY	2641	GGAGTGTCCGTCGCTGACCAACAGCATCTCTGTCTATCATTTCTTTTGGCAATACT	2700
QY	1561	CCACCTGAAAACCTGTAGTGAAGTACAGCACTGCATCTGCTGTAGCCAGTAGCCAC	1620	Db	2668	ACAGTGTCTCGAGCGGTACCACTCAGCATCTGTTTCCATCATATCTCTCGGGAATACT	2727
Db	1588	CCCCGGGAAAACCTGTAGTGAAGTTCACACTGTCATCTGTTCCGTAGCCAGTACCAC	1647	QY	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGTGTATCATATGAAGAGGGTCTTACCAGG	2760
QY	1621	TCAGTAGGACTTCTAATGTTTGCAGAGAAAATAATGATATATTTTGGCATCTCTGACCTT	1680	Db	2728	TTAATGAGTATGACCAATGCAACTTTTATTTGTGTATCATATGAAGAGGGCCCGCCAGG	2787
Db	1648	TCTGTAGGCTCTTAAGTCTGCGAGAGAAAATGCAATCATGTTGGCGTCTCGCCACCTC	1707	QY	2761	CCACCTTAGAACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT	2820
QY	1681	TTTCCCTATTCAGTAATCAATGAGGCTTCTGTGATGATTTACTGTTGGTGGGCTGTCA	1740	Db	2788	CGGCTTAGACCTGGCAGCCCGACACCTCTCTAAGGGGAGGGATTCCTCTCAGGCTCCAGT	2847
Db	1708	TTTCCCTATTCAGTATGATGTTGCGAGATGGAACACTGGTGGCTGGATCGCAATGCGA	1767	QY	2821	AATATTGTGCAAGGACAGATTAACAGTTGTGCACTGTGCTGCTTCCCTCGGCTCGAT	2880
QY	1741	GATGGTCTGTGTACCTGTGGCAAAATGGAATCTGGTGCATTTGGATCGTTGTGTGATGGG	1800	Db	2848	AACTTTGTGCAAGGACAGATTAACAGCCGCTGCGCTGTGCTGCTGCTGCTGCTGCTGCTG	2907
Db	1768	GACGGCTCTGTGTATGTTGCGAGATGGAACACTGGTGGCTGGATCGGATCGTGTGCAATGGGA	1827	QY	2881	GCTGATCACTCTGGCTCTGACCTCTCTCTGCTCTGCTCTTACATACTGTTTCTTAGTA	2940
QY	1801	ATAACAGAGTGTGAGATTTTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTTGATTC	1860	Db	2908	GCTGATCACTCTGGCTCTGACCTCTGCTCTCTGCTCTGCTCTGCTCTGCTGCTGCTGCTG	2964
Db	1828	ATCACAGCAGTGGAGATACCTGAACGCTTGTGACGAAGCCGCTTCTCGCGCTGTAGACTCT	1887	QY	2941	AATGAAGGTTGGAGTCAGTTAGCTGTATGCTGTGTTATGCTGTGACAGCTTACTGGGA	3000
QY	1861	CTTAGTCATCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGCTGCTCTT	1920				


```
Db 2965 AATGAAGCTGGAGCCAGTTAGCGCCATGCATCTGTTATGTTGCTGACCTGCTGGGG 3024
QY 3001 TTGATAAATTTAGGCTCCCTTTCTGGAGATGCTCGCCGCGAAGATGCGAAGATCGATGC 3060
Db 3025 CTGGAGAGATTAGGCTCTCTCTCTGGAGATGCTAGCTCGAAGATGCGCAGGACCGATGC 3084
QY 3061 TTGGAGGTGAGAGAGCCGACAGAGCCCTGCTTCTGGCGGAACTGAGAGAAATTGAGCAG 3120
Db 3085 TTGGAGGTGAGAGAGCTGACAGGCGCTGCTTCTAGCTGAGCTGAGAGAGATTGAGCAG 3144
QY 3121 GCAGGCGAGGAAGGACCATTTGATGCTGGGCTCTTACTTACTCAGTACATAGACCAAC 3180
Db 3145 GCAGGCGAGGAAGGAGAGATGACACCTGGGCTCTTACTTACTCAGTATAGACCAAT 3204
QY 3181 GTCATATCACTGGAGTCAATCAGAAAGCGGCGACATATCACACCGCTCTCTGATGCC 3240
Db 3205 GTCATATCACTGGGCTCACGGCGAAGCCATGACACTATGGCAGCTGCTCCAGATGCC 3264
QY 3241 TCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGAGATGACCTTGTTCACGATGACATCAC 3300
Db 3265 TCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGAGATGACCTTGTTCAGTATGATGACATCA 3324
QY 3301 ACTGGTGTCTTCAAGTGTCCACAAATGAAAAAATTTCTACATCTTTACGAGGAAGA 3360
Db 3325 GCAGGTGTCTTCAAGTGTCCACAGATGAAGAAAAATCTCCAGCTTTACGAGGAAGA 3384
QY 3361 CGGAAGCAAGTACCGCTATTGTTTACTTGGAGTAATAGGAGCTGAATTTGGTGTGAA 3420
Db 3385 AGGAGCAGGCCACGGCTATTGTTCTCTGGAGTGATAGGGGCGAGGTTTGGAGCTGAA 3444
QY 3421 ATTGAACCTCTTAACATTTGACAGCACTCGAGCTTAGCCAAATTTCTGAGGATTC 3480
Db 3445 ATTGAACCCCGAAGCTGTGACCAAGCTCGAAGCTCCAGTCAAAATCCCTGAAGCTTT 3504
QY 3481 GGGTGACTAGTGGTGATCCAACTACTCGTGGCCAGACATCTTGCAGAGCACTGACG 3540
Db 3505 GGCCTGACAGCGGAGGCTCCAACTACTCCCTGGCCAGACACACGTGCAAGGCACTGACA 3564
QY 3541 TTTCTTCTGCTACAGCTCCAAAGCCCAAACTTCTCCACAGCAGCACTATCCGAGAAACA 3600
Db 3565 TATCTCTGCTACAGCCACCAAGTCCCAAGCTCCCTCTCATGACACCATCCGGAGACT 3624
QY 3601 GCATATGATCTGANTGGAAGTGGGTTCACTGTTTGGAGGCTTTACATGATGTGCGGT 3660
Db 3625 GCCACTGACTGATTTGGGCGAGGTTCACTGTATGGAGCCCTTACATGATGTGTCGCT 3684
QY 3661 GTTCTGATGGGCTTCTGAACTTTGTGCGGATGCGGAGAAACAATTCGCCAACATCA 3720
Db 3685 GTCTGTATGGGCTGCTGAGCTCTGTGAGACGCTGAGAAACAATTCGCCAACATCA 3744
QY 3721 ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCCGCTCTGCGAGGCAATGCCCTCTCG 3780
Db 3745 ATGGGGTCTGCTCTGAGCCCTGAGCTGACTCTGCGGCTCCGCAAGGCAAGCCCTGCT 3804
QY 3781 CTCTATGCAAGCCGACAGACCAAGCCCTTTCATCAACCATATGACCAAGAGGTATCACAGA 3840
Db 3805 CTCTATGCAAGCCGACAGACCAAGCCCTTTCATCAACCATATGACCAAGAGGTATCACAGA 3864
QY 3841 CATACGGCTCTTGCAGCAAAATACCAATACAGCAGAAATATGCAACACAACTCTTGA 3900
Db 3865 CACACGGGCTTTCAGCGCAACACCCAGTCCGAGCAGAGATATACACACCAACCCCTGGGG 3924
QY 3901 CGAGCTAAAGGGGAAATTTGAGAGTCAATGAAATCTTATTTGAAAGATGCCCACAGAT 3960
Db 3925 AGGGCTAAAGGGGAAATCTGAGAGTCAATGAAATCTTATTTGAAAGATGCCCAGGAT 3984
QY 3961 GTTGTGGATCTTCTGAGGAGGTATGGAATCATATGATGATGCTGCTGAGGATCTTTA 4020
Db 3985 GTGTGGATCTTCTGAGGAGGTATGGAATCATATGATGATGCTGCTGAGGATCTTTA 4044
QY 4021 GTTAAAAAGAGGCTTCAAGAAATGTTTCCAGCAGCATCTGAGGTTCTACATGTCAGC 4080
Db 4045 GTGAAAAAGAGGCTTTCAGGAGTGTGTTTCCAGCTATCTGAGGTTCTACATGTCAGC 4104
```

```
QY 4081 TATTATGAGCGGAATACAGAAATAGCAGTTTGAGCTTGCATGTTCACTGGCCCTGTATC 4140
Db 4105 TATTATGAGCGCAGTCACAGAAATCGCAGTTGGAGCAGCCCATGGCTCAGTAGCCCTGTAT 4164
QY 4141 GACATCCGAGCTGGAATAATGTCCAGAAATCCATGACACAAAGGACCAATCACTGCAGTG 4200
Db 4165 GACATCCGAGCTGGAATAATGTCCAGAAATCCATGAGGACCAATCACTGCAGTG 4224
QY 4201 GCTTTTGCTCTGATGGAATATCTTGCACCTACTCAAAACACTGACAGCAACATTTCT 4260
Db 4225 TCCTTTGCTCTGATGAGCGTTTACCTTGCACCTACTCGAAACACGACAGCCACATTTCC 4284
QY 4261 TTTTGGCAGATGAACACGTCACCTGCTGCGAAGCATGCGCATGCTGAACCTCGGACCTCAG 4320
Db 4285 TTTTGGCAGATGAACACCTCCTCTGGAAGCATTTGAATGCTGAACCTCAGCTCCTCAG 4344
QY 4321 CTGCGCTGCATTAACCTTACCAGGTGCCCTGTGCGAGCCGCTGCGCCCGCTCCAC 4380
Db 4345 CTGCGCTGCATTAACCTTACCAGGTGCCCTGTGCGAGCCGCTGCGCCCGCTCCAC 4404
QY 4381 AATGCCCTCAAGCTGGCCGCTCATCTGGAATTCACACCGCAAGCTCATCTCATGGCC 4440
Db 4405 AAGGCCCTTAAGTTGGCCGCTCATCTGGAATTCACACCGCAAGCTCATCTCATGGCC 4464
QY 4441 CATGACGGAGGAGGACCGCTTTCATGCTTAA 4473
Db 4465 CATGATGGAGGAACACCGCTTTCATGCTTGA 4497
```

RESULT 5
AF188124

LOCUS Mus musculus TGF-beta resistance-associated protein TRAG (Trag) 4467 bp mRNA linear ROD 15-SEP-2001

DEFINITION mRNA, partial cds.

ACCESSION AF188124

VERSION AF188124.1 GI:15624072

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Sanders, S. and Thorgeirsson, S.S.

TITLE TRAG: a novel gene associated with TGF-beta resistance

JOURNAL Unpublished

REFERENCE

2 (bases 1 to 4467)

AUTHORS Sanders, S. and Thorgeirsson, S.S.

TITLE Direct Submission

JOURNAL Submitted (20-SEP-1999) LEC, NIH, 37 Convent Drive, Bldg. 37, Rm

3C28, Bethesda, MD 20892-4255, USA

FEATURES

Location/Qualifiers

1..4467

/organism="Mus musculus"

/mol_type="mRNA"

/strain="129SvJ"

/db_xref="taxon:10090"

/chromosome="18"

/map="18D.1-E.3"

/tissue_type="brain"

1..>4467

/gene="Trag"

1..>4467

/gene="Trag"

/codon_start=1

/product="TGF-beta resistance-associated protein TRAG"

/protein_id="AAL03982.1"

/db_xref="GI:15624072"

/translation="MAGNSLVLPVLMGRKAPTHCISILLTDDGGTIVTGDHGGIC

LWDVSELVNPRALLFGHTASITCLSKACAGDKRYTVSASANGEMCLWDVNDGRCI

EFTKLACTHTGIGQFYQFVGNGQGRLLCHGYPEILVVDATSLVLSVSKISPDW

ISSMSIHSQTDQEDTVALLKVMIVTSEMSGMQDTEPIFEESKPIYCCNCQ

SISFCAPTQSRLLVVCISKYWRVFDAGDYSLLCSGPFSENGQTWTGDFVSNADKVIITWE

Qy	1801	ATAACAGCAGTGTGAGATTCTTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTGATTC	1860
Db	1801	ATCACAGCAGTGTGAGATCTGAACGCTTGTGACGAAGCCGTTCTGCGCTGTAGACTCT	1860
Qy	1861	CTTAGTCTATCCAGCAGTCAACCTTAAACACAGCTATGACGAGAGCTAGTCTTGTGCTGCTT	1920
Db	1861	CTTAGTCTATCAGCAGTCAACCTGAAACACAGCCATGACAGACGGAGTCTCGCTGCCCCTT	1920
Qy	1921	AAAAATATGGCCCATCATAAAGCTACAAACCCCTTGCAACTAACTCTTTGGCTCTCTGAGGCA	1980
Db	1921	AAAAATATGGCCCAACCATAGCTGCAAAACCCCTTGCAAGAACCTTTTGGCTTCTGAGGCC	1980
Qy	1981	TCTGACAAAGGGAATTTATCTTAAATATTCATAACTCTCCTGATGGTTCAAGCAATAAAG	2040
Db	1981	TCTGACAAAGGGAATTTATCTTAAATATTTCCATAACTCTCCTGATGGTTCAAGCAATAAAG	2040
Qy	2041	ACAAACCTTAACAGACCCGGGACATACATGTCTATTTCTTGATGTGGAAGCGTTTGAATATT	2100
Db	2041	ACAAACCTTAACAGACCCGGGATATCCATGTGCTTTTCTTGATGTGGAAGCTTTTGAATATT	2100
Qy	2101	CAACTCTCTGACTGGAAGAGCCTCTAGGCCGAATACTGCTCTTATATTTCCCCAGAGAAATTG	2160
Db	2101	CAACTCTCTGACTGGAAGAGCCTCTAGGCCGAATACTGCACTTATTTCCCCAGAGAACTG	2160
Qy	2161	CAAAAAAGCATCTGCGAGTTCCAGACAAAGGGGGCTCTTTTTTAACTGGAACACGAGCAGCA	2220
Db	2161	CAGAAAGCATCTGCGAGTTCCAGACAAAGGGGGCTCTTTCTTAACCTGGAAGCGAGCAGCG	2220
Qy	2221	GTTCCTCTCCAAACAGTGTAAAGAAACCATCAAGAGAGAAATCAAGGAAACCTCTCTTGAT	2280
Db	2221	GTTCCTCTTTCAGCAAGTGTAAAGAAACATATCAAGAGAGAAATCAAGGAAACCTCTCTGGAT	2280
Qy	2281	GATGAAGAGGAGATGAGGAGATTAATCAGGCAGAGAGGAAGAAAGTGAATCTGTAATAT	2340
Db	2281	GAGGAGGAGAGAGGAGGAGGCGAGGAGGCGAGGAGGAGATGATGATGACCTTGAGTAC	2340
Qy	2341	CGGTCCAGCAAAATCAAAAGCCATTGACCCCTATTAGAATATTAATTTAACTATGACACTGCA	2400
Db	2341	CGGGCCAGCAAGTCCAAGCCACTGACCCCTACTAGAATACAACCTTACTATGGGATACGCA	2400
Qy	2401	AAGCTGTTTATGCTCTGCTCTTCCAGCTCTGGGTTTGAATGAAGTACTGGGATGAAGTTTGC	2460
Db	2401	AAATTAATTCATGCTCTGCTCTCCAGCTCTGGGGCTTGAATGAAGTTCTGGATGAAGTTTGT	2460
Qy	2461	CTGATCGCTTGGAAATGCTGAAACCCCACTGCACCGTATCTGTTGGCCCTCTTGTCAAGA	2520
Db	2461	CTGATCGCTTCGGAATGCTGAAACCCGCACTGCACCGTGTCTTTGGTCTCTCTATCGAGA	2520
Qy	2521	GGAGGCCATATGTCATGATGCTGCCGGGTTTAAATCAGCGCTGCTGTGTAACCTGTCCACAT	2580
Db	2521	GGAGGTCATATGTGCTGATGCTTCCCGGTTTAAATCAGGCTGCTGGAAGAACTGTCTCAT	2580
Qy	2581	GGGAAACACAGATGAGGAAGAGCTGCCAGCGTCTGAGGGAGTAGGAAAGGGAACCTTAC	2640
Db	2581	GCGAAACAGAAAGTAGGAAGAAAGCTGCCAGCAGCGGAGCGGCTAGGCAAGGGAACCTAC	2640
Qy	2641	GGAGTGTCCGTGCGTCCACACACAGCATCTCTGTCTATCATTTCTTTTGGCAATACT	2700
Db	2641	ACAGTGTCTCGAGGGGTCAACATCAGCATCTGTGTGTCATCATATCTCTTGGCGAATACT	2700
Qy	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAGGGTCTCTACGAGG	2760
Db	2701	TTAATGAGTATGACCAATGCACTTTCAATTTGGCGATCACAATGAAGAGGGCCCCACGAGG	2760
Qy	2761	CCACTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT	2820
Db	2761	CCGCCTAGACCTGGCACCCAGACCTCTCTAAGGCGAGGGATTTCCCTCCAGGCTCCAGT	2820
Qy	2821	AATATTGTCAGAGCAGATTTAAACAGTTTGTGCGACCTGTGCTTTTCGGCTCCGCTCTGAT	2880
Db	2821	AACATTGTGACAGGACAGNTTAAACAAAGCCGCTGCGCTGTGCTTTCTGCTCGGCTCTGAC	2880

QY	2881	GCTGATCACTCTGGCTCTGACCCCTCTCTCTGCTCTCTCTTACATACCTGTTCTTTAGTA	2949
DB	2881	GCTGATCACTCTGGCTCTGACTCTGCGCT--CTCCTGCTTTTACCTACTCTTTCTTTAGTA	2937
QY	2941	AATCAAGGTTGGAGTTCAGTTAGCTGCTATGCACCTGTGCTTATGCTGCCAGACCTACTTGGGA	3000
DB	2938	AATGAAGCTTGGAGCCAGTTTAGCCGCATGCACTGTGTATGTGTGCTGACCTGCTGGG	2997
QY	3001	TTGGATAAATTTAGGCCCTCCCTCTTCTGGAGATGCTGGCCGGAAGATGGCAAGATTCGATGC	3060
DB	2998	CTGGAGAGATTCAGGSCCTCTCTCTCTGGAGATGCTAGCTCGAAGATGCGAGGACCGATGC	3057
QY	3061	TTGAGGTGAGAGAAGCGGACAGGCCCTGCTTCTTGGCGGAACTGAGAGAATTTGAGCAG	3120
DB	3058	TTGAGGTGAGAGAGGCTGACAGGCCCTGCTTCTAGCTGAGCTGAGAGGATTTGAGCAG	3117
QY	3121	GCAGCGAGGAAGAACCAATTGATGCTGGGCTCTTACTTACTCTAGTACATAGACCA	3180
DB	3118	GCAGACGGAAGGAGACGATCGACACTGGGCTCTTACTTACTCTAGTATATGAGCCAT	3177
QY	3181	GTCCATATCACTGGAGTTCACATCAGAAAGCGCGCAGACTATACCAAGGCTCTCTGATGCC	3240
DB	3178	GTCCATATCACTGGCGTTCACGGCGGAGCCATGCAAGACTATGGCAGCTGCTCCAGATGCC	3237
QY	3241	TCAGGGCTCGAAGCAAAAGTTCAGAGGAGAAAGATGACCTTGTGTGACGATGACATCACC	3300
DB	3238	TCGGGCCCGAAGCCAAAGTTCAGGAAGAAGAGCATGACCTGGTGGATGATGACATCACT	3297
QY	3301	ACTGGTTGCTTATCAAGTGTCCCACAANTGAAAAAATTTCTACNTCTTACGAGGAAGA	3360
DB	3298	GCAGGTTGCTTGTCAAGTGTCCCAAGATGAAAAAATCTCCACGCTTTCAGGAAGAAAGA	3357
QY	3361	CGGAAGCAAGCTACCGCTATTGTTTACTTGGAGTAAATAGCAGCTGAATTTGGTGTGAA	3420
DB	3358	AGGAAGCAGGCCACGGCTATTGTTCTCTGGAGTGTAGGGGAGAGTTTGGAGCTGAA	3417
QY	3421	ATTGAACCTCTCTAAACTATTTGACACAGACCTCGAAGCTCTAGCCAAATTCCTGAGGGATTC	3480
DB	3418	ATTGAACCGCGCAAGCTGCTGACAGGCTCGAAGCTCCAGTCAAAATCCCTGAAGGCTTT	3477
QY	3481	GGGTTGACTAGTGTGATCCAACTACTGCTGGCCAGACATCTTTCGACAGGCACCTGACG	3540
DB	3478	GGCTTGACCAAGCGAGGCTCCAACTACTCTCTGGCCAGACACAGTGCAGAGGCATGTACA	3537
QY	3541	TTTCTTCTGCTCAGACCTCCAAGCCCAAACTTCTCCACACAGCACTATCCCAAGAAACA	3600
DB	3538	TATCTTCTGCTCAGCCACCAAGTCCCAAGCTCCCTCTCATAGACCACTCCCGAGAACT	3597
QY	3601	GCCATTCATCTGATTTGACGTGGGTTCACTGTTTGGAGCCTTACATGAGATGTGCGCT	3660
DB	3598	GCCACTGACCTGATTTGGGCGAGGGTTCACTGTATGGAGCCCTACATGGATGTGTCGCT	3657
QY	3661	GTTCTGATGGGCTTCTCGAACTTTGTGCCGATGCGGAGAAACAACTTGCCCAACATCACA	3720
DB	3658	GTCCTGATGGGCTGCTGGAGCTCTGTGTCAGACGCTGAGAAAACAATTCGCAACATCACA	3717
QY	3721	ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCGCTCTCGAGGCAATGCCCTCTCG	3780
DB	3718	ATGGGGCTGCTCTGAGCCCTGCACTGACTCTGCGGCTCTCGAAGGCAAGCCCTGTCT	3777
QY	3781	CTCATTCGCCAGCCGACACCCGCTTCCATCCACCAATAGCCAAAGGATGACACAGA	3840
DB	3778	CTCATGCCACAGCCAGACCAACCGCTTCCATCCACCAATAGCTAAGGAGGTCCACAGA	3837
QY	3841	CATACGGCTCTTGAGCGAAATACCAATCAGCAGAGATATGACACAACTCTTGCA	3900
DB	3838	CACACGGCCCTTGACGCGAAACGCCAGTCCAGCAGAGATATACACCAACCACTCTGGCG	3897
QY	3901	CGAGCTAAAGGGGAAATTTTTCAGAGTCAATTGAAATCTTATTCGAAAGATGCCCAAGAT	3960
DB	3898	AGGGCTAAAGGGGAAATCTCTGAGAGTCAATTGAAATCTCATTTGAAAGATGCCCAAGAT	3957
QY	3961	GTGTGTGATCTTCTCGTGGAGGTTATGAGACATCAATTATGTACTGCTTGAAGATCTTTA	4020

QY 601 ACTGAGCCAAATTTTGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAGCATC 660
Db 601 ACTGAGCCAAATTTTGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAGCCTC 660
QY 661 TCTTTTGTGCAATTTACACAAAGTCACATTTTGTGTGTGTGTTCCAAATATTTGAGGGTG 720
Db 661 TCTTTTGTGCAATTTACACAGAGTCGCTCTTGTGTGTGTGTTCCAAATATTTGAGGGTG 720
QY 721 TTGATGCGGAGACTATTCCTTGTGTGTGTGTTGAGGTCTAGTGAATGACAGACATGG 780
Db 721 TTGATGTCGGCACTACTCCTCTCTGTGTTGAGGTCTAGTGAATGACAGACATGG 780
QY 781 ACCGGGGGACTTTGTCTCATCAGATTAAGTCAATCTTTGGACAGAAATGGCAAAAT 840
Db 781 ACTGGAGGGACTTTGTGTCTGACAGAAAGTCAATTTTGGACTGAATACGGGACAGAT 840
QY 841 TATATTTACAACTACCTGCCAGTTGCTTCCAGCTAGTGATTCATTTCCGAGTGATGGT 900
Db 841 TACATTTACAACTCCCTGCCAGTTGCTTCCAGCTAGTGATTCATTTCCGAGTGAGTG 900
QY 901 GGGAGGCGAGTTGAAATTTAAATTCCTCTGTGTACAAATATTCCTTTGGATCGAAAGAT 960
Db 901 GGGAAAGCAGTTGAAATCTGATTCCTCCGTGCGAGCATAGCCCTCTTGGATCAGAGAT 960
QY 961 AAGAGTTGCTAATTTGCTCCTCTGTTACTCGGTTCTTCTATGATCGACAGATATTTTC 1020
Db 961 AGAGAGTTGGAATTTGCTCCTCTGTTACTCGGTTTCTTATGATCGAAGGAATATTTG 1020
QY 1021 CATAAACTGTTAAATTCAGGGTGATTTCTCTGGAAGTTGAAATTTTGGAAATATCAGAC 1080
Db 1021 CATAACTGTTAAATTCAGGGTGATTTCTCTGGAAGTTGAAATTTTGGAAATATCAGAC 1080
QY 1081 ACAGCTGATAAACAGGAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTGCA 1140
Db 1081 ATAGCAGACAAACAGGAAGCAATGAAGGCTGAAAGCAAGCACTTGTATTAGTTGCA 1140
QY 1141 GAGGCAATTTGATTAACCTGTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1141 GATGCAATTTGACAACTGAAAGCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACCAGCACATGACGA 1260
Db 1201 CCCAAGCAGCAACGAACTTAAAGTAACTGCGAGTGTCTACATACCAGCACAGCGGCG 1260
QY 1261 CTTGTTTGTGTCGTGGAAGTGAAGCATAGTTATTTGTAAGTGTGCTGCTGCTGCTGCTGCT 1320
Db 1261 CTTGTTTGTGCGCGGGAAGCAGGAAGCATCATTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 GTACAGCTGTTGCAAGGGGAACAATGCTCAGAAAGAGTTGGCCACCTCAGACAACTC 1380
Db 1321 GTTCAGCTGTTGCAAGGGGAACAATGCTCAGAAAGAGTTGGCCCGCCGCAAGAACCCCTC 1380
QY 1381 CGTGTCTATCGGAACCAAGTCAATGTTGCTATATCTCATCAGGTCTCAGCTCGGTAT 1440
Db 1381 CGTGCCACCGGAACCAAGTCAATGTTGCTATATCTCATCAGGTCTCAGCTCGGTAT 1440
QY 1441 GATCAAAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCAATATTTGGGACATATTT 1500
Db 1441 GATCAAGATACCTGATATCOGAGGTGTGGATTTTTCGTCATCATTTGGGACATTTT 1500
QY 1501 TCTGGAGAAATGAACATATCTTCTGTGTTCAATGGTGTGAGATTAATCAACTCTAGATT 1560
Db 1501 TCCGGAGAAATGAACATATCTTCTGTGTTCAATGGTGTGAGATTAATCAACTCTCTGTC 1560
QY 1561 CCACCTGAACCTGTAGTGAAGTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1561 CCGCAGAAATCTGTAGTGAAGATTTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCAATATTTGGCATCTCTGTCACCTT 1680
Db 1621 TCTGTAGGCTGCTAAGTCTGCGAGAGAAATGCAATATGCTGCGGCTCTCTGTCACCTG 1680
QY 1681 TTTCTTATTCAGTAATCAAAATGGAGGCCCTTCTGATGATTAACCTGCTGCTGCTGCTGCT 1740

Db 1681 TTCCCTATTTCAGGTGATCAAGTGGAGCCCTTTCGACGACTACTCTGGTGGTGGGTGCACG 1740
QY 1741 GATGGTTCTGTGTACGTCTGGCAAAATGGATACCTGGTGCTTGGATCGTTGTGTGATGGG 1800
Db 1741 GACGGCTCTGTGTGTGTGTCGAGATGGACACACTGGTGGTGGACCGCTGTGCAATGGG 1800
QY 1801 ATAAACAGCAGTGTGAGATTTAAACGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
Db 1801 ATAAACAGCCTGTGAGATTTCTCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
QY 1861 CTTTGTGTCATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGCTGTGTGTGTGTGTGT 1920
Db 1861 CTGAGTCAACCCAGCAGTCAACCTTGAAGCAAGCATGACACGGGGAGTCTGCGCGCCCTT 1920
QY 1921 AAAAATATGCCCCCATCATTAAGCTTACAAACCCCTTTCGCAACTAACCTTCTGGCTTCTGAGGCA 1980
Db 1921 AAAACATGCCCCCACCACAGCTGCAAAACCCCTTTCGCAACTAACCTTCTGGCTTCTGAGGCC 1980
QY 1981 TCTGACAGGGAAATTTTAACTTAAATATTCATAAATTCCTAACTCCCTGTGTGTGTGTGTGT 2040
Db 1981 TCTGACAGGGAAATTTTAACTTAAATATTCATAAATTCCTAACTCCCTGTGTGTGTGTGTGT 2040
QY 2041 ACAAACCTAACACAGCCGAGACATACATGTGTATTTCTTGTGTGTGTGTGTGTGTGTGT 2100
Db 2041 ACAAACCTAACTGACCCGGATATCCATGTGTCTTTTGTGTGTGTGTGTGTGTGTGTGT 2100
QY 2101 CAACTCTGACTCAAGAGCCCTCTAGGCCGAATACCTGTCTTATTTTCCCGAGAGAAATTTG 2160
Db 2101 CAACTCTGACTCAAGAGCCCTCTAGGCCGAATACCTGTGTGTGTGTGTGTGTGTGTGTGT 2160
QY 2161 CAAAAAGCATCTGGCAGTTTCAGAAAGGGGCTCTTTTAACTGGAAGAACGAGCAGCA 2220
Db 2161 CAGAAAGCATCTGGCAGTTTCAGAAAGGGGCTCTTTTAACTGGAAGAACGAGCAGCA 2220
QY 2221 GTTCTCTTCCAACTGAAAGAAACGATCAAGAGAAACATCAAGNAACCTCTCTTGAT 2280
Db 2221 GTTCTCTTCCAACTGAAAGAAACGATCAAGAGAAACATCAAGNAACCTCTCTTGAT 2280
QY 2281 GATGAAGAGGAGATCAGGAGATAATGAGCAGAGAGGAGGAGAAAGTGTATCTGGAATAT 2340
Db 2281 GAGGAGAGGAGCAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2340
QY 2341 CGGTCCAGCAAAATCAAAAGCCATTTGACCTTATTAAGAAATATAATTTAACTATGGAACCTGCA 2400
Db 2341 CGGGCCAGCAAGTCCAAACCACTTACCTACTAGAAATCAACCTTACTATGGAATCCGCA 2400
QY 2401 AAGCTGTTTATGTCTGCTTCCAGCTGCGGCTTGAATGAAATGTAATGTAATGTAATGTA 2460
Db 2401 AAATTAATCATGTCTCTGCTCCACGCTGCGGCTTGAATGAAATGTAATGTAATGTAATG 2460
QY 2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGCAACCGTATCGTGTGCTCTTGTCAAGA 2520
Db 2461 CTGATCGCTCGGCAATGCTGAAACCACTGCAACCGTATCGTGTGCTCTTGTCAAGA 2520
QY 2521 GAGGCCATATGTCTACTGTGCTGCGGGTTAATCAGCCTCTTGTAACTGTCAACAT 2580
Db 2521 GAGGTCTATGTCTCTTGTGCTTCTGCTTAACTAGCTCTGGAAGCTACTGCGAG 2580
QY 2581 GGGAAAACAGAGTGAAGAGAGCTGCGCGCTGTGAGGGAGTAGGAAGGAACTTTAC 2640
Db 2581 GCGAAAACAGAGTGAAGAGAGCTGCGCGCTGTGAGGGAGTAGGAAGGAGGAGGAGGAGG 2640
QY 2641 GAGGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
Db 2641 AAGGTCTCGGAGCGGTCAACCGCAATCTGTTGTCATCATTTCTTGTGCAAAATACT 2700
QY 2701 TTAATGAGTATGCACTAATGCACTTTTATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2760
Db 2701 TTAATGAGTATGCACTAATGCACTTTTATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2760
QY 2761 CCACTTAGACCAAGCAACCCCGAGCCTTTCTAAGGCAAGGGGTTCCCTCTCCAACTTCCAGT 2820

Db 2761 CCGCTAGACGAGCCACCCAGACCTTTCTAAGCAAGGATTCCTCCCCAGCCTCCAGT 2820
QY 2821 AATATTGTGCAAGCAGATTAACAAGTTGCTGCACTGTCGTTTCGGCTCGGTCTGAT 2880
Db 2821 AACAATTGTGCAAGCAGATTAACAAGTCTGCGCCTGTC---TCTGCTCGGTCTGCC 2877
QY 2881 GCTGATCACTCTGGCTCTGACCTCTCTGCTGCTCTGCTTTTACATACCTGTTTCTTAGTA 2940
Db 2878 GCCGACACTCTGGCTCTGGCTCTGGCT---CTCTGCTTTAGTACCTGCTTTTAGTG 2934
QY 2941 AATGAAGTTGGAGTCAAGTTAGTCTGCTATGCACTGTTTATGCTGCCAGACCTACTGCGGA 3000
Db 2935 AATGAAGTGGAGCCAACTAGCTGCTGCACTGCTGTCATGCTGCCGAGCTGCTGGGG 2994
QY 3001 TTGGATTAATTTAGGCTCTCCCTCTTGGAGATGCTGGCCGCAAGATGGCATGCTC 3060
Db 2995 CTGGGTAAATTCAGGCTCTCTTCTGGAGATGCTAGCTCGAAGATGGCATGCTGCTC 3054
QY 3061 TTGGAGGTGAGAGAGCGGCACAGGCCCTGCTTCTGGCGGAACTGAGAAGAAATTGAGCAG 3120
Db 3055 TTGGAGGTGAGAGAGGCTGCACAGGCCCTTCTTCTAGCAGAGCTGAGAAGAAATTGAGCAG 3114
QY 3121 GCAGGCGAGGAGGAGCCATTTGATGCTGGGCTCTTTACTTACTCTCAGTACATGAGACCAC 3180
Db 3115 GCAGGCGGAGGAGACTATTTGATACCTGGGCTCTTTACTTACTCTCAGTACATGAGACCAC 3174
QY 3181 GTCATATCACTGGAGTCAATCAGAAGCCGCGGAGACTATACACGAGCTCTGATGCC 3240
Db 3175 GTCATATCACTGGAGTCAAGCGGAGGAGCCATGAGACTATGTCAGCTGCTCAGATGCC 3234
QY 3241 TCAGGGCTCAAGCAAAAGTCCAGGAGGAGAGAGATGACCTGTTGACGATGACATCACC 3300
Db 3235 TCGGGGCGAGAGCCAAAGTCCAGGAGGAGAGAGATGACCTGTCGAGCGATGACATCACC 3294
QY 3301 ACTGGTTGCTTATCAAGTGTCCACAAATGAAGAAATTTCTACATCTTACGAGGAAAGA 3360
Db 3295 ACTGGTTGCTTATCAAGTGTCCACAAATGAAGAAATGTCACATCTTACGAGAAAGA 3354
QY 3361 CGGAGCAAGCTACCGCTATTTGTTTACTTGGAGTATAGGAGCTGAATTTGGTGTGTA 3420
Db 3355 AGGAAGCAGGCCACTGCTATTTGTTCTCTGGAGTGTAGGAGCAGAGTTTGGAGCTGAA 3414
QY 3421 ATTGAACCTCTTAACTATTGACGAGACTCGAAGCTCTAGCCAAATTCCTGAGGAGTTC 3480
Db 3415 ATTGAACCAAAACTGCTGACGAGACTCGGAGCTCTAGTCAAAATCTCTGAGAGATT 3474
QY 3481 GGGTTGACTAGTGTGGATCAACTACTCTGCTGCCAGACATATTTGCAAGGCACTGACG 3540
Db 3475 GGTTTGACAAGTGGAGTTCCAACTACTCTGSCCAGACATAGGTGCAAGGCACTGACA 3534
QY 3541 TTTCTTCTGCTACAGCTCCAGCCCAAACTTCTTCCACAGACATATCCGAGAGACA 3600
Db 3535 TTTCTTCTGCTACAGCCCAAACTTCCAAACTTCTCTCTATGACCATCCGAGAACT 3594
QY 3601 GCCATTGATCTGATTGACGCTGGTTCACGTTTGGAGCTTACATGAGATGTCGCT 3660
Db 3595 GCCATTGACCTGATCGGGCGAGGGTTTACCCTGTGGAGCTTACATGAGAGCTGCTGCT 3654
QY 3661 GTTCTGATGGGCTTCTCGAACTTTGTCCGATGCCGAGAAACAACTTGCCCAACATCACA 3720
Db 3655 GTCTGATGGGCTGCTGGAGCTGTGTGAGATGCTGAGAAGCAGCTGGCCCAACATCACA 3714
QY 3721 ATGGGGTTGCTCTGAGCCGACGAGCTGATCTCGGCCGCTCTGCGGAGGATGCCCTCTCG 3780
Db 3715 ATGGGGTGTGCTCTGAGCCCTTGCAGCTGACTCTGCCCGATCCGCAAGACACGCCCTTTCT 3774
QY 3781 CTCAATGCCACCGCAGACCAACCGCTTCTATACACCACTAGTCCAAAGAGGTACACAGA 3840
Db 3775 CTCAATGCCACCGCAGACCAACCGCTTCTATACACCACTAGTCCAAAGAGGTGACAGA 3834
QY 3841 CATACGGCTCTTGAGCAAAATACCAATACAGCAGAAATATGACACAACTCTTGCA 3900
Db 3835 CACAGGCGCTTGGCAGCAAAATACCACTGCCAGCAGATATCCACACCAACCACTGGCA 3894

QY 3901 CGAGCTAAAGGGGAAATTTTGAGAGTCAATTGAAATTTCTTATTGAAAAGATGCCACAGAT 3960
Db 3895 AGGCTAAAGGGGAAATCTGAGAGTCAATTGAAATTTCTTATCGAAAAGATGCCACAGAT 3954
QY 3961 GTTGTGATCTTCTCGTGGAGTTATGGACATCATPATATGATCTGCTTGAAGATCTTTA 4020
Db 3955 GTTGTGATCTTCTTGTGGAGTCTATGAGACATCATGATCTGCTTGAAGATCTTTA 4014
QY 4021 GTTAAAAAGAAAGTCTTCAAGAAATGTTTCCAGCCATCTGAGGTTCTACATGGTCAGC 4080
Db 4015 GTTAAAGAAAGAGGCTTCTCAGGAGTGTTCAGAGCCATCTGAGGTTCTACATGGTCAGC 4074
QY 4081 TATTATGACGGAAATCACAGAATAGCAGTTGAGCTCGCCATGGTTCTAGTGGCCCTGTAC 4140
Db 4075 TATTATGACGGAGTCAAGAAATGCAAGTTGAGGACGCGCATGGCTCAGTGGCCCTGTAT 4134
QY 4141 GACATCCGAGCTGGAAATGTGAGCAATTCATGAGACACAGAGGACCAATCATCTGAGTG 4200
Db 4135 GACATCCGAGCTGGGAAATGTGAGCAATTCAGGACACAGAGGACCTATCATCTGAGTG 4194
QY 4201 GCTTTGCTCTGATGGAAAGATATCTTGCCACTACTCTAAACACTGACAGCCACATTTCT 4260
Db 4195 TCTTTGCTCTGATGGGCTTACCTTGCACCTACTCTAAACACTGACAGCCACATTTCT 4254
QY 4261 TTTTGGCAGATGAACACGTCACCTGCTGGGAAGCATCGGCATGCTGAATCTGGCACCCTCAG 4320
Db 4255 TTTGCGCAGATGAACACCTCCTCTTGGGAAGCATGGCATGCTGAATCTGACCATCTCAG 4314
QY 4321 CTGCGCTGATTAACACCTTACAGAGTGCCTCTGTGACGCGCGCTGTCAGCCGCTCCAC 4380
Db 4315 CTGCGCTGATCAAGACCTTACAGGTACCTCAGTGCAGCCGCTATCCCTGGCTGCCAC 4374
QY 4381 AATGCCCTCAGCTGCCCGGCTCATCTGAGCTTCCAAAGCAAGTCACTCTCATCTGAGCC 4440
Db 4375 AAGGCCCTCAGTGTGGCGGCTCATCTGAGCTTCCAAAGCAAGTCACTCTCATCTGAGCC 4434
QY 4441 CATGACGGGAAGGAGCAGCCGCTTCATGCTCTAA 4473
Db 4435 CAGATGGGAAGGAGCAGCCGCTTCATGCTGA 4467

RESULT 7

AX244638
LOCUS AX244638 4494 bp DNA linear PAT 28-SBP-2001
DEFINITION Sequence 1 from Patent WO0166739.
ACCESSION AX244638
VERSION AX244638.1 GI:15859532
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Thorgeirsson,S. and Sanders,S.
TITLE [dna encoding a tag gene (tgf- g(b) resistance associated gene)
and its protein product
Patent: WO 0166739-A 1 13-SEP-2001;
JOURNAL THE DEPARTMENT OF HEALTH & HUMAN SERVICES (US)
FEATURES
source
1. .4494
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
/notes="Rat Brain"
28. .4494
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC88574.1"
/db_xref="GI:15859532"
/translation="MAGNSLIVLPVWGRKAPTHCISILLTDDGTVTGCHDGOIC
LMDLSELEVNPRALLFGHTAAITCLSKACAGDQKQYTVSANGEMCLWDVNDGRCI
EFTKLACTHTGIQYQFSVGNQREGLCHGHPYELVVDATSLVLSKISPDW
ISSMSIIRSHRTQEDTVVALSVTGILKMWIVTSEISGLQDTEPIFEESKPYCQNCQ

SLSPCAFTORSLLVCSKVRVPDAGDYSLSCSPSEDDGQWTGDFVSADKVLWTE
NQSYIYKLPASCLPASDSFRSDVGRKAVENLIPVQSHSLDQKDRLEUVIPPTRFY
GCKYLKHLILQSSSGLSIWNJADIQOENEGUKTTICLSQDAFDKJPCPAG
IIDQSVIPNSNEKVTASYIIPAHGLVCGREDGSIIVPATQTAIVQLGQEHML
RRWPBHRTLLRHNRKVTCLLYPHQVSARYDQRYLISGVDFSVIIMWIFSGEMKHF
CVHGETITQLLVPPENCARVQHCVSASVSDHSLSLREKCKIMLASRHLPIQVI
KWRPSDYLVVGGCTDGSVCVWMDTGALDRCAMGITAVEILINACDEAPRAVDLSHP
AVNLKQWTRSRSLAALKMAHKLQTLATNLLASBDSKGNLKYSHNSLMVQAIKTN
LTDPIHVPDVEALIIQLLTHEASRPNTALISPENLKASGSSDKGSLFLTKRMA
VLFQVKEITKENIKHLLDEEEMRQREESPEYRASKEPLTLLEYNLTMW
TAKLFMSCLHAGKNEVDEVDLRLGKLPCHTVSFLSLRGHMSMLPGYNQAG
KLQKAEACRKGATPESVGVTVSVRAVTOHLSIIISLANTLMSMTNATFTGDHM
KGPTRPPECTDLSKARDSPASSNIVOGLOKQAAAPVSAASADHSSASASPAL
RCPFLVNEGSHQALAMCVMLPDLGLGKRPFLLEMLARWQDRCLUEVREAAQULL
ABURIEQAGRKTITWAPYLQYMDHVISPGVTAEAMQMAAPDASPEAKQJBE
EHLVDVDDITITGLSSVPOMKMSSTYERKQATATVLGVIGAEFGABIEPPKLLT
RPRSSQIPBGFGLTSGSNYSLARHTKALTPLLQPPSPKLPPIHRTIRTAIDLIG
RGFTWEPYMDVSAVLMGLLELCADEKOLANITWGLPLSPAADSARSARHLSLIT
ARPPAFITTIKAEVHRHTALANTOSQSLHTTTLAEAKGEILLRVIIEKMPDQV
DLIVVMDIIMYCLEGLSVKKGLOECFPAICRFYVWSYIYERSHRIIAGARHSVALY
DIRTKCQITIHGKGPITAVSPAPDGRYLATYSNTDSHISFWQNTSLSLGSIQMLNSA
POLRCIKTYQVPPVQPASPGSHNLRRLRLIWTNSRNVILMAHDGKEHPRMV"

ORIGIN

Query Match 79.1%; Score 3537.8; DB 6; Length 4494;

Best Local Similarity 87.2%; Pred. No. 0;

Matches 3901; Conservative 2; Mismatches 564; Indels 6; Gaps 2;

Qy	1	ATGCGAGAAACAGCCTTGTTCTACCCATTTGTTCTTTGGGTGCAAAAGCGCCACACAT	60
Db	28	ATGCGAGAAACAGCTAGTTCTGCCATTTGTTCTTTGGGCGCAAGCACCCACAC	87
Qy	61	TGCATCTCAGCGGPACTTTTAAACAGATGATGGGCGCACGATCGTAAACAGATGTCACGAC	120
Db	88	TGTATTTGCTCAATCTGCTGACAGATGATGGGCGCACGATTTGACTGGATGCCAGAT	147
Qy	121	GGCAAAATATGCTCTGGGATCTTTAGTAGAGTGCAGATGCAAAATTAATCTCTGAGCACTGTTG	180
Db	148	GGCAGATATGCTCTGGGATCTCTCAGAGAGCTGGAAGTTAATCTCCCGAGCACATTA	207
Qy	181	TTTGGTCATACAGCATCAATCAGTTGTTGTCTAAAGCTTGCTTCCAGTGACAAACAG	240
Db	208	TTTGGTCACAGCAGCCATCATTGTTGTCTAAAGCCTGTGCTTCTGGAGACAAACAG	267
Qy	241	TATATTTGATGTCATCTGAAGTGGAGAGATGTGCTCTGGAGTGTGAGTGTGAGGAGA	300
Db	268	TACACCGTGAGTGCCTCGCAAAATGGAGAGATGTGCTCTGGAGTGTGAATGAGCGCAGA	327
Qy	301	TGTATTTGAATTTACAAATTTAGCTTGACACATCTGGCATACAGTTCTACAGTTCTCT	360
Db	328	TGTATTTGAATTTACAAATTTAGCTTGACACATCTGGCATACAGTTCTACAGTTCTCT	387
Qy	361	GTTGGGAATCAGCGAGAGAGAGCTTTTATGCCAGGACATTAACCTGAAATCCTTGTT	420
Db	388	GTTGGGAATCAGCGAGAGAGAGCTTTCTGTGCCATGGGATTAACCTGAAATCCTGTT	447
Qy	421	GTGGATGCTACAGCCTTTGAAGTATTAATCTCTCTTAGTATCAAGATATACCCAGACTGG	480
Db	448	GTGGATGCCACAGCCTTTGAGTGTGTTATCTCTTGGTATCGAAGATCTCTCCAGACTGG	507
Qy	481	ATTAGCTCCATGAGTATTTTGCATCCCGAGACACAGAGGACACAGTGTGTAGCATCTC	540
Db	508	ATCAGCTCCATGAGTATTTATCTCTCTCAGCGGACACAGAGGACACAGTGTGTAGCATCTC	567
Qy	541	TCGGTAGCTGCGATCTCGAAGTCTGAGTTGTTACCTCGGAAATAGTGACATACGAGGAT	600
Db	568	TCCGTGACAGGATTTCTGGAAGGTGTGATTTGTGACCTCTGAAATTTAGTGATTTGACGAC	627
Qy	601	ACTGAGCCAAATTTTGAAGGAGGATCCAAACCAATTTTATTTGTGAGATTTGCCAAAGCATC	660
Db	628	ACTGAGCCAAATTTTGAAGGAGGATCCAAACCAATTTTATTTGTGAGATTTGCCAAAGCCTC	687
Qy	661	TCTTTTGTGCAATTTACAAAGGTGCACTTTTGGTGTGTGTTTCCAAATATTTGAGGGGTG	720

Db	688	TCITTTTGTGCAATTCACAGAGAGTCTCTTGTGTTGTATGCTCAAAGTACTGGAGGGTG	747
Qy	721	TTGATGCCGAGAGATTAATCTTGTGTTGTTGAGTCTCTAGTGAATAATGACACACATGG	780
Db	748	TTGATGCTGGGACTACTCTCCCTGCTGTGTTGAGTCTCTAGTGAAGATGACACACATGG	807
Qy	781	ACGCGGGGGAGACTTTGTCTCATCAGATAAAGTCAATCAATTTGGACAGAAAAATGGGCAAGT	840
Db	808	ACTGGAGGGAGACTTTGTGTCTGCAGACAAAGTCAATTAATTTGGACTGAAAAACGGCAGAGT	867
Qy	841	TATATTTTACAAACTTACCTGCGAGTTCCTCCAGCTAGTCAATTTCCGCGAGTGACGTG	900
Db	868	TACATTTTACAAACTCTCCCTGCGAGTTCCTCCAGCTAGTCAATTTCCGCGAGTGACGTG	927
Qy	901	GGGAAGGCACTTGAAAAATTTAAATTCCTCTCTGTGTAACAAATATCTCTCTTGGATTCGAAAAAGAT	960
Db	928	GGGAAGGCACTTGAAAAATCTGATTCCTCCGCTGCAGCATAGCTCTTTGGATCAGAGGAT	987
Qy	961	AAAGAGTTGCTAAATTTGTCTCTCTGTTAATCGGTTCTTCTATGATGACAGAGAAATATTTTC	1020
Db	988	AGAGAGTTGCTAAATTTGTCTCTCTGTTAATCGGTTCTTCTATGATGACAGAGAAATATTTTC	1047
Qy	1021	CATAAATGTTAATTTCAAGGTTGATTTCTCTGGAAGTTGAATATTTTCGAAACATATCAGAC	1080
Db	1048	CATAAATGTTAATTTCAAGGTTGATTTCTCTGGAAGTTGAATATTTTCGAAACATATCAGAC	1107
Qy	1081	ACAGCTGATATAACAGGGAAGTGAAGAGGCTGCAATGACAACTTCTTATTTAGTTTGCAC	1140
Db	1108	ATAGCAGACAAAACAGGAGCCATGAAGGCTTAAGAGCAACTTGTATTTAGTTTGCAC	1167
Qy	1141	GAGGCAATTTGATATAAATGAAATCTCTGCTGCTGGAATTAATGATCAGCTGAGTGTGAT	1200
Db	1168	GATGCAATTTGACAAAATGAAAGCTTGTCTGCTGGAATTAATGATCAGCTGAGTGTGAT	1227
Qy	1201	CCCAATAGTAAATGAACTCTTAAAGTAACTGCAAGTGTGTACATACCAGACACATGGACGA	1260
Db	1228	CCCAACAGCAACGAAACCACTTAAAGTAACTGCAAGTGTGTACATACCAGACACATGGACGA	1287
Qy	1261	CTTGTGTTGCTGCTGAAAGATGGAAGCATAGTTATTTGTACTGCTGCCACACAGACGGCCATA	1320
Db	1288	CTTGTGTTGCTGCTGCGGAGAGAGCATCAATTTATGCTGCTGCCACACAGACGGCCATA	1347
Qy	1321	GTCAGCTGTTGCAAGGGGAAACATGCTCAGAAAGAGTTGGCCACTCAACAGAACATCTC	1380
Db	1348	GTTTCACTGCTGCGGAGAGAAACATGCTCAGACGAGTTGGCCCGCCACAGAACCTCTC	1407
Qy	1381	CGTGGTCATCGGAAACAAAGTCAATGTTTGTCTATATCTCATCAGGTCTCAGCTCGGTAT	1440
Db	1408	CGTGGTCATCGGAAACAAAGTCAATGTTTGTCTATATCTCATCAGGTCTCAGCTCGGTAT	1467
Qy	1441	GATCAAAAGATACCTGATATCTGAGAGTGTGGATTTTTCAGTCAATAATTTGGGACATATTT	1500
Db	1468	GATCAAAAGATACCTGATATCTGAGAGTGTGGATTTTTCAGTCAATAATTTGGGACATATTT	1527
Qy	1501	TCGGGAAATGAAACATATCTTCTGTGTTTCATGTTGGTGTGAGATTAATCAACTTCTAGTT	1560
Db	1528	TCGGGAAATGAAACATATCTTCTGTGTTTCATGTTGGTGTGAGATTAATCAACTTCTAGTT	1587
Qy	1561	CCACCTGAAACATGTTAGTGCAGAGTACAGCATGTCATGCTCTGTAGCCAGTACAC	1620
Db	1588	CCGCCAGAAACATGTTAGTGCAGAGTTCACATCTGGCTGTGTTCTGTGGCCAGTACAC	1647
Qy	1621	TCACTAGGACTTCTTAAAGTTTGGCAGAGAAAAATGCAATATTTGGGATCTCTCGTCACTT	1680
Db	1648	TCCTGAGGGCTGTAGTGTGCGAGAGAAAAATGCAATATTTGGGATCTCTCGTCACTT	1707
Qy	1681	TTTCTTATTAAGTAATCAAAATGAGGCGCTTCTGTGATTAATTTACCTGTTGGTGGGTGTTC	1740
Db	1708	TTTCTTATTAAGTAATCAAAATGAGGCGCTTCTGTGATTAATTTACCTGTTGGTGGGTGTTC	1767
Qy	1741	GATGGTTCTGTGATGCTGCGAAATGGAATATCTGGTGCATTTGGATCTGTTGTGTGATGGG	1800

1768 GACGGCTCTGTGTGTCTGGCAGATGGACACTGGTGGCTGGACCGCTGTGCAATGGGG 1827
1801 ATAAAGCAGTTGAGATTTCTAAACGCTTGTGTGATGAAGCTGTCTCGTGTCTGTGATTC 1860
1828 ATAAAGCAGTTGAGATTTCTCAATGTCTGTGACGAAGCTGTCTCCCTGACGAGTGGACTCA 1887
1861 CTTAGTCATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGTGTCTTGTGCTCTTT 1920
1888 CTGAGTCCAGCAGTCAACCTTAAACAGCTATGACGAGAGTGTCTTGTGCTCTTT 1947
1921 AAAAATATGCCCATCATAGCTTACAAACCTTTGCAACTAACTCTTTGGCTTCTGAGGCA 1980
1948 AAAACATGSCCCACACAGCTTGCAACCTTTGCAACTAACCTTTTGGCTTCTGAGGCC 2007
1981 TCTGAACAAGGGAATTTACTTAATATTTCTATAACTCTCCCTGATGTTTCAAGCAATAAAG 2040
2008 TCTGAACAAGGGAATTTACTTAATATTTCTATAACTCTCCCTGATGTTTCAAGCAATAAAG 2067
2041 ACAACCTTACAGCCGGACATACATGTCTATTTCTTGTGATGGAAGGTTGATTTAT 2100
2068 ACAACCTTAACTGACCCGGATATCCATGTGCTTTTCTTGTGATGGAAGCTTTGATTTAT 2127
2101 CAACTCTGACTGAAGAAGCTCTAGGCCGAATACTGCTCTTATTTTCCCAAGAAATTTG 2160
2128 CAACTCTGACTGAGAGGCTCTAGGCCGAATACTGCACTTATTTTCCCAAGAAATCTG 2187
2161 CAAAAAGCATCTGGCAGTTGAGCAAAAGGGGGCTCTTTTAACTGCGAAACAGAGCA 2220
2188 CAGAAAGCATCTGGCAGTTGAGCAAAAGGGGGCTCTTTCTGACTCGAAACAGAGGCA 2247
2221 GTTCTCTTCCAAAGTGAAGAAACGATCAGAGAAACATCAAGAAACATCTCTTTGAT 2280
2248 GTTCTTTTCCAGCAAGTGAAGAAACATCAAGAGAAACATCAAGAGGAGCACTCTTTGAT 2307
2281 GATGAAGAGAGGATGAGAGATAATGAGGCAGAGAGGAAGAAAGTGAATCTCTGAATAT 2340
2308 GAGGAGGAGGACGAGGAGAGGATGATGAGGCAGAGGAGGGAAGAAAGTGAATCTGATAC 2367
2341 CGGTCCAGCAAAATCAAGCAATTTGACCTTATTAAGATATTAATTAATTAATTAATTAAT 2400
2368 CGGCGCAGCAAGTCCAAACCACTTACCTTACTAGAATACAACTTACTATATGAGATACGCA 2427
2401 AAGCTGTTATGTCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTTGAATGAAGTTTGC 2460
2428 AATATTTATGATGCTGCTCTCAAGCTGGGGTTTGAATGAAGTCTGGAATGAAGTTTGC 2487
2461 CTGGATCGCTTGGAAATGCTGAACCCCACTGACAGCTATCGTTTGGCTCTTTGTCAAGA 2520
2488 CTGATCGCTCTGGCATGCTGAAACCACTGACAGCTGTCTTTGGTCTCTCTATCCAGA 2547
2521 GGAGGCGATATGCACTGATGCTGCGGGTTTAAATCAGCTGCTGTTGTAACCTGTCAAT 2580
2548 GGAGGTCAATATGCTTGTATGCTTCTGTTTATAATCAGGCTGCTGGAAGCTACTGCAG 2607
2581 GGGAAACCAAGTATGAAGGAGCTGCGAGCTCTGAGGAGTATGAGAGGGAACCTTAC 2640
2608 GCGAAAGCAGAGCAGGACGGAAGGGCCAGCAACGAGAGCGTAGGCAAGGGGACCTAC 2667
2641 GGAGTGTCCGCTGCGCTCACACACAGCATCTCTGTCTATCATTTTCTTTGGCAAAATFAC 2700
2668 ACAGTGTCCGAGGGGTCAACACGCAATCTGTGTGTCATCATCTCTGCGCAAAATFAC 2727
2701 TTAATGATATGACCAATGCAATTTTATTTGTTGATCATATGAAGAGGGTCTCTACAGG 2760
2728 TTAATGAGCATGACCAATGCGAGTTTCAATTTGGAGATCACATGAAGAGGGGCCCAACAGG 2787
2761 CCACTTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCCACTTCCAGT 2820
2788 CCGCTTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGGATTTCCCTCCCACTTCCAGT 2847
2821 AATATTTGCAAGGACAGATTAACCAAGTTGTGCACTCTGTCTGTTTCCGCTCGGTCTGAT 2880
2848 AACATTGTGCAAGGACAGATTAACCAAGTGTCTGCGCTGTGTC---TCTGCTCGGTCTGCC 2904

2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTGCTCTTTTACATACCTGTTTCTTAGTA 2940
2905 GCGACCACTCTGGCTCTGCCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2961
2941 AATGAAGTGTGGAGTCAAGTCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2962 AATGAAGTGTGGAGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3021
3001 TTGATATAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCGGAAGATGGAAGATCATGCTG 3060
3022 CTGGGTAAATTTAGGCTCTCCCTTTCTGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3081
3061 TTGAGGTGAGAGAGCCGACAGGCCCTGCTCTTGGCCGAAGTCTGAGAGAAATTTAGGACAG 3120
3082 TTGAGGTGAGAGAGCTGCAAGGCCCTTCTTCTAGCAGAGCTGAGAGAAATTTAGGACAG 3141
3121 GCAGGCGAGAGAGGAGCCATTTGATGCTGGGCTCTTACTTACTCTCAGTACATAGACCCAC 3180
3142 GCAGGCGAGAGGAGGACTATTTGATGCTGGGCTCTTACTTACTCTCAGTACATAGGACCAT 3201
3181 GTCATATCACTTGGAGTCAATCAAGAGCCGCGAGACTATCAACAGGCTCTGATGCC 3240
3202 GTCATATCACTTGGAGTCAAGGCGGAAGCCATGCGAGCTATGCGAGCTGCTCCAGATGCC 3261
3241 TCAGGGCTGAAGCAAAAGTCCAGGAGGAAGAGCATGACCTTGTTCAGATGACATCAC 3300
3262 TCGGGGCCAAGAGCCAAAGTCCAGGAGAGAGAGTACCTCTGTCAGCATGACATCAC 3321
3301 ACTGTTGTCTTATCAAGTGTCCCAAAATGAAAAAATTTTACATCTTTACGAGGAAAGA 3360
3322 ACTGTTGTCTTATCAAGTGTCCCAAAATGAAAAAGATGTCACATCTTTACGAGAAAGA 3381
3361 CGGAAGCAAGCTTACCGCTATTTGTTTACTTGGAGTAAATGAGGCTGAAATTTGCTGTGAA 3420
3382 AGGAAGCAGGCCACTCTCTATTTGTTCTCTGGGAGTATAGGAGCAGAGTTTGGAGCTGAA 3441
3421 ATTGAACCTCTTAAACTTATTTGACAGACTCGAAGCTCTAGCCAAATTTCTGAGGAGATTC 3480
3442 ATTGAACCAACCAAACTGCTGACAGACTCGGAGCTCTAGTCAAAATTTCTGAGAGATTT 3501
3481 GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATATTTGCAAGGCACTGACG 3540
3502 GGTTCGACAGTGGAGGTTCCAACTACTCTCTGGCCAGACATACGTCGCAAGGCACTGACA 3561
3541 TTTCTTCTGCTACAGCTCCAGCCCAAACTTCTCTCCACAGCAGCATATCCCAAGACA 3600
3562 TTTCTTCTGCTACAGCCCAAGTCCCAAACTTCTCTCTCATAGCACTATCCCGAGAACT 3621
3601 GCCATTTGATCTGATTTGAGTGGGTTTCACTGTTTGGGAGCTTTACATGATGTTGTCGCT 3660
3622 GCCATTTGATCTGATTTGAGGAGGTTTCACTGTTGAGGAGCTTTACATGAGAGTGTCTGCT 3681
3661 GTTCTGATGGGGCTTCTCGAACTTTGTGCGGATGCGCGAGAAACAACTTGCACAACTACCA 3720
3682 GTTCTGATGGGGCTGCTGGAGCTGTGTCAGATGCTGAGAGCAGCTGGGCCAACATCACA 3741
3721 ATGGGGTGGCTCTGAGCCAGCAGCTGACTGGCCGCTCTGCGAGGAGTGGCTCTGCTG 3780
3742 ATGGGGTGGCTCTGAGCCCTGAGCTGACTCTGCGCCGATCTGCGCCGATCTGCGCCGATCT 3801
3781 CTCATTTGCCACCGCCAGAGCCAGCCGCTTTCATCAACCACTATAGCCAAAGAGGTACACAGA 3840
3802 CTCATAGCACCCGCCAGAGCCAGCCGCTTTCATCAACCACTATAGCCAAAGAGGTGACAGA 3861
3841 CATACGGCTCTTGGCAGCAATATCCCAATCAAGCAGAGATATGACACAACTTGTGCA 3900
3862 CACACGGCCCTTGCAGCAAAATATCCAGTCCAGCAGAGTATCCACACCACTACCACTGCGCA 3921
3901 CGAGCTTAAAGGGGAAATTTTGGAGTCAATGAAATTTCTTATTTGAAAGAGTCCCAAGAT 3960
3922 AGGGCTTAAAGGGGAAATCTCTGAGAGTCAATTTGAAATTTCTTATTCGAAAAAGATGCTTACG 3981

QY	2226	CTTCCAAAGTGAAGAAACGATCAAGAGAACATCAAGGAACACCTCCTTGATGATGA	2285	QY	3306	TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAGACGGAA	3365
Db	1377	CTTCCAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCCTTGATGATGA	1436	Db	2457	TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAGACGGAA	2516
QY	2286	AGAGGAGGATGAGGAGATATAGGCGAGAGAGGAAAGTATCTCGTAATATCGGTG	2345	QY	3366	GCAAGCTACCGCTATTGTTTTTACTTTGGAGTAATAGGAGCTGAATTTTGGTGTGAAATTTGA	3425
Db	1437	AGAGGAGGATGAGGAGATATAGGCGAGAGAGGAAAGTATCTCGTAATATCGGTG	1496	Db	2517	GCAAGCTACCGCTATTGTTTTTACTTTGGAGTAATAGGAGCTGAATTTTGGTGTGAAATTTGA	2576
QY	2346	CAGCAAAATCAAGCCATTGACCCCTATTAGAAATATAATTTAACTATGAGACACTGCAAAAGCT	2405	QY	3426	ACCTCCTAAACTATTGACACAGACCTGGAAGCTCTAGCCAAATTCCTGAGGGATTCGGGTT	3485
Db	1497	CAGCAAAATCAAGCCATTGACCCCTATTAGAAATATAATTTAACTATGAGACACTGCAAAAGCT	1556	Db	2577	ACCTCCTAAACTATTGACACAGACCTGGAAGCTCTAGCCAAATTCCTGAGGGATTCGGGTT	2636
QY	2406	GTTTTATGCTCCTGCTTCAAGCCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTGGCTGGA	2465	QY	3486	GACTAGTGTGATCCAACTACTCGCTGGCCAGACATATCTTGCAAGGCCACTGACGTTTTCT	3545
Db	1557	GTTTTATGCTCCTGCTTCAAGCCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTGGCTGGA	1616	Db	2637	GACTAGTGTGATCCAACTACTCGCTGGCCAGACATATCTTGCAAGGCCACTGACGTTTTCT	2696
QY	2466	TCGCCCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGGCTCTTGTCAAGAGGAGG	2525	QY	3546	TCTGCTACAGCTTCCAAGCCCCCAAATCTCCTCCACACAGCATATCCGAAGAACACGCCAT	3605
Db	1617	TCGCCCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGGCTCTTGTCAAGAGGAGG	1676	Db	2697	TCTGCTACAGCTTCCAAGCCCCCAAATCTCCTCCACACAGCATATCCGAAGAACACGCCAT	2756
QY	2526	CCATATGTCACTGATGCTGCGGGTTTATAATCAGCCCTGCTTGTAAACTGTFCACATGGGAA	2585	QY	3606	TGATCTGATTGACGTGGGTTCACTGTTTGGGAGCCTTACATGGATGTCTCGGCTGTTCT	3665
Db	1677	CCATATGTCACTGATGCTGCGGGTTTATAATCAGCCCTGCTTGTAAACTGTFCACATGGGAA	1736	Db	2757	TGATCTGATTGACGTGGGTTCACTGTTTGGGAGCCTTACATGGATGTCTCGGCTGTTCT	2816
QY	2586	AACAGAAATAGGAAGAGCTGCACCGTCTGAGGAGTAGGAAGAGAACTTACCGAGT	2645	QY	3666	GATGGGGCTTCTCGAACTTTTGTGCCGATGCCGAGAAACAACTTGCCCAAATCAACAATGGG	3725
Db	1737	AACAGAAATAGGAAGAGCTGCACCGTCTGAGGAGTAGGAAGAGAACTTACCGAGT	1796	Db	2817	GATGGGGCTTCTCGAACTTTTGTGCCGATGCCGAGAAACAACTTGCCCAAATCAACAATGGG	2876
QY	2646	GTCCGCTGGCGTCAACACAGCATCTCTGCTATCATTTTCTTGGCAATACTTTAAT	2705	QY	3726	GTTTCCCTCTGAGCCACAGCAGCTGACTGCGCCCGCTCTGCGAGGCATGCGCTCTCGCTCAT	3785
Db	1797	GTCCGCTGGCGTCAACACAGCATCTCTGCTATCATTTTCTTGGCAATACTTTAAT	1856	Db	2877	GTTTCCCTCTGAGCCACAGCAGCTGACTGCGCCCGCTCTGCGAGGCATGCGCTCTCGCTCAT	2936
QY	2706	GAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGGTCTTACAGGCCACC	2765	QY	3786	TGCCACCGCAGACACCGCCCTTCACTACACCACTAGCCAAAGAGGTACACAGACATAC	3845
Db	1857	GAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGGTCTTACAGGCCACC	1916	Db	2937	TGCCACCGCAGACACCGCCCTTCACTACACCACTAGCCAAAGAGGTACACAGACATAC	2996
QY	2766	TAGACCMAGCACCCGAGACCTTCTAAGGCAAGGGTTCCTTCCCACTTCCAGTAATAT	2825	QY	3846	GGCTCTTGCAGCAAAATACCAATCACAGCAGAAATATGCAACAACTCTTTCGACGAGC	3905
Db	1917	TAGACCMAGCACCCGAGACCTTCTAAGGCAAGGGTTCCTTCCCACTTCCAGTAATAT	1976	Db	2997	GGCTCTTGCAGCAAAATACCAATCACAGCAGAAATATGCAACAACTCTTTCGACGAGC	3056
QY	2826	TGTGCAAGGACAGATTAACAGTTGCTGCACCTGCTGTTTCCGCTCGGTCTCATGCTGA	2885	QY	3906	TAAAGGGGAAATTTTGAGAGTCAATTGAAATTTCTTATTTGAAAAAGATGCCACAGATGTTGT	3965
Db	1977	TGTGCAAGGACAGATTAACAGTTGCTGCACCTGCTGTTTCCGCTCGGTCTCATGCTGA	2036	Db	3057	TAAAGGGGAAATTTTGAGAGTCAATTGAAATTTCTTATTTGAAAAAGATGCCACAGATGTTGT	3116
QY	2886	TCACTCTGGCTCTGACCCCTCTGCTCTGCTCTGCTTTTACATACCTGTTTCTAGTAATGA	2945	QY	3966	GGATCTTCTCGTGGAGGTTATGGACATCATTTATGTACTGCTTGAAGGATCTTTAGTTAA	4025
Db	2037	TCACTCTGGCTCTGACCCCTCTGCTCTGCTCTGCTTTTACATACCTGTTTCTAGTAATGA	2096	Db	3117	GGATCTTCTCGTGGAGGTTATGGACATCATTTATGTACTGCTTGAAGGATCTTTAGTTAA	3176
QY	2946	AGTTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGGGATTTGA	3005	QY	4026	AAAGAAAGGTCTTTCAAGAAATGTTTCCAGCCATCTGCAGGTTCTACATGTCAGCTATTAA	4085
Db	2097	AGTTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGGGATTTGA	2156	Db	3177	AAAGAAAGGTCTTTCAAGAAATGTTTCCAGCCATCTGCAGGTTCTACATGTCAGCTATTAA	3236
QY	3006	TAAATTTAGGCTCCCTCTCTGAGATGCTGGCCCGAAGATGCAAGATCGATGCTTGA	3065	QY	4086	TGACCGGAATCAAGAAATAGCAGTTGGAGTCTGCCCATTTGTTTCAGTGGCCCTGTACGACAT	4145
Db	2157	TAAATTTAGGCTCCCTCTCTGAGATGCTGGCCCGAAGATGCAAGATCGATGCTTGA	2216	Db	3237	TGACCGGAATCAAGAAATAGCAGTTGGAGTCTGCCCATTTGTTTCAGTGGCCCTGTACGACAT	3296
QY	3066	GGTGAGAGAACCCGACAGGCCCTTCTTGGCGGAACTGAGAGAAATGAGCAGGCGAGG	3125	QY	4146	COGGACTGGAATAATGTCAGACATCCATGGAACACAGGGGACCAATCACTGCACTGGCTTT	4205
Db	2217	GGTGAGAGAACCCGACAGGCCCTTCTTGGCGGAACTGAGAGAAATGAGCAGGCGAGG	2276	Db	3297	COGGACTGGAATAATGTCAGACATCCATGGAACACAGGGGACCAATCACTGCACTGGCTTT	3356
QY	3126	CAGGAAGGAAGCCATTGATGCTGGGCTCCTTACTTACCTCAGTACATAGACACCGTCAT	3185	QY	4206	TGCTCTGATGGAAGATATCTTGCACACTACTCAAACTGACAGGCAATTTCTTTTTTG	4265
Db	2277	CAGGAAGGAAGCCATTGATGCTGGGCTCCTTACTTACCTCAGTACATAGACACCGTCAT	2336	Db	3357	TGCTCTGATGGAAGATATCTTGCACACTACTCAAACTGACAGGCAATTTCTTTTTTG	3416
QY	3186	ATCACCTGGAGTCACATCAGAGCCGCGCAGACTATCACACCGCTCCTGATGCTCAGG	3245	QY	4266	GCAGATGAACACTCACTGCTGGGAAGCATGCGCATGCTGAACTCGGCACTCAGCTGCG	4325
Db	2337	ATCACCTGGAGTCACATCAGAGCCGCGCAGACTATCACACCGCTCCTGATGCTCAGG	2396	Db	3417	GCAGATGAACACTCACTGCTGGGAAGCATGCGCATGCTGAACTCGGCACTCAGCTGCG	3476
QY	3246	GCCTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTTTGTTGACGATGACATCACCACTGG	3305	QY	4326	CTGCATTAAACCTACAGGTGCCCTTGTGAGCGCGGTCCCGGGCTCCCAACATGC	4385
Db	2397	GCCTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTTTGTTGACGATGACATCACCACTGG	2456	Db	3477	CTGCATTAAACCTACAGGTGCCCTTGTGAGCGCGGTCCCGGGCTCCCAACATGC	3536
				QY	4386	CCTCAAGCT-GGCCCGGCTCATCTGAGCTTCCAAACCGCAACGTCATCTCATGGCCCATG	4444

||||| 3537 CCTCAAGCTGGCCCGGCTCATCTGGACTTCCACCGCAACGTCATCTCATGGCCCATG 3596
||||| 4445 ACGGGAAGGACCGCTTTCATGCTCTAA 4473
||||| 3597 ACGGGAAGGACCGCTTTCATGCTCTAA 3625

RESULT 9
AB011113
LOCUS Homo sapiens mRNA for KIAA0541 protein, partial cds. PRI 10-APR-1998
DEFINITION
ACCESSION AB011113
VERSION AB011113.1 GI:3043605
KEYWORDS KIAA0541 protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H.,
Nomura, N., and Ohara, O.
Prediction of the coding sequences of unidentified human genes. IX.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro
DNA Res. 5 (1), 31-39 (1998)
98290545
9628581
2 (bases 1 to 6072)
Ohara, O., Nagase, T. and Ishikawa, K.
Direct Submission
Submitted (13-FEB-1998) Otsu Ohara, Kazusa DNA Research Institute,
DNA Technology, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
[E-mail: cdmainfo@kazusa.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914]
Location/Qualifiers
1. 6072
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG4294"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1. 6072
/gene="KIAA0541"
1. 3485
/gene="KIAA0541"
/codon_start=3
/product="KIAA0541 protein"
/protein_id="BAA25467.1"
/db_xref="GI:3043606"
/translation="REFYQREVFKLLIQDSSGELNINWISDTADKQGSBEGLAMT
TSISLQEAEDKLNPCAGIIDLQSVIPNSNEPLKVTSVYIPAHGLVCGREDGSIVI
VPATQAI VQLQGHMLRGPHTLRGRNKVTCILYPHQVCSYSDORVILSGGV
DFSVIITDFSGEMHIFCVHGGEITQLLPVPPENCARSARVHCISVADSHVGLLSLR
EKKICMLASRHLFPPIQVIRKPSDDYLVVCGSDGVVWMDTGALDRVCVMTASVRI
LNACDEAPVAAVDLSHPAIVNLQAMQTRSLAALKNWHLKQLTLATNLASEADPKG
NLPKYSHNSLWQAIKTNLTDPIHVLFFDVEALIIQLTEASRPWTALISPENLQK
AGSSDKGSLPTGKAAVLPQVKTIIKENIKHLDDDEDEEIMRORRESDPEY
RSKSKPLTLELITLMDTAKFMSCLHAWGLNEVLDDVCLDRGLMKLPHCTVSGLL
LANTLMSMTNATPIGDMKKGPTRPPPTPDLSEGVGKGTGVSRVATTOHLLSIIS
VSRSDADSHGPPSPAPALHTCFVNEGWSOLAAMHCVMLPDLGLDKFRPPLLEML
ARWQDRCLVREAAQALLAELELRLEQAGKEAIDAWAPYLPQYIDHVI SPGVTS
AQIITTPASGPEAKVQEEHLDVDDITTCGLSSVPQMKKISTISEERRKQATV
LGVICAEFGAELEPPKLLTRPSSQIPEGFLTSGGNSYLARHTKALFTLLQIP
PSKLPPTHTRTADLIGRFTWEPYMDVSAVLMGLLELCAEAKOLANI TWGLP
LSPADASARSHALSLIATAPPAFTITAKEVHRHTLAANTOSQONHTTLARA
KGEILAVIELLEKMPDVVLLVEVMDIIMVLEGLSVKKGLQECFPAICRFYVVS
YVERNRIAVGARHGSVALYDINTGKQTIHGHKFTITAVAPDQGRFATINTSDSH
ISFQWNTSLLSIGMNSAPQLRCIKTYQVPPVQSPGSHNALKARLIWTSNRNV
ILMAHDGKEHRFMV"

ORIGIN

Query Match 77.9%; Score 3485; DB 9; Length 6072;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 989 CTCGGTCTCTATGATGCAGAGAAATATTTCCATAAATGTTAAATTCAGGGTGATTCCT 1048
DB 1 CTCGGTCTCTATGATGCAGAGAAATATTTCCATAAATGTTAAATTCAGGGTGATTCCT 60

QY 1049 CTCGAAGTGTGAATATTTGGAAACATATCAGACACAGCTGATAAAGAGGGAAGTGAAGAAG 1108
DB 61 CTGGAAGTGTGAATATTTGGAAACATATCAGACACAGCTGATAAAGAGGGAAGTGAAGAAG 120

QY 1109 GGTGCAATGACAACTTCTATTAGTTTCAAGAGGCAATTTGATAAATCTGAATCTCTTGTG 1168
DB 121 GGTGCAATGACAACTTCTATTAGTTTCAAGAGGCAATTTGATAAATCTGAATCTCTTGTG 180

QY 1169 CTGCTGGAATATAGATCAGCTGAGTGTGATTTCCCAATAGTAAATGAACCTCTTAAAGTAA 1228
DB 181 CTGCTGGAATATAGATCAGCTGAGTGTGATTTCCCAATAGTAAATGAACCTCTTAAAGTAA 240

QY 1229 CTCGAAGTGTGATACATACAGACACATGAGGAGCTTGTGTGCTCGTGAAGATGGAAGCA 1288
DB 241 CTCGAAGTGTGATACATACAGACACATGAGGAGCTTGTGTGCTCGTGAAGATGGAAGCA 300

QY 1289 TAGTTATTGTACCTGCCACACAGACGCGCCATAGTACAGCTGTTGCAAGGGGAACACATGC 1348
DB 301 TAGTTATTGTACCTGCCACACAGACGCGCCATAGTACAGCTGTTGCAAGGGGAACACATGC 360

QY 1349 TCAGAAAGGTTGGCCACCTCAGAGAAACATCTCGTGGTCAATCGGAACAAAGTCAATGTT 1408
DB 361 TCAGAAAGGTTGGCCACCTCAGAGAAACATCTCGTGGTCAATCGGAACAAAGTCAATGTT 420

QY 1409 TGCATATCTCATCAGTCTCAGCTCGTGTATGATCAAGATACCTGATATCTGGAGGTG 1468
DB 421 TGCATATCTCATCAGTCTCAGCTCGTGTATGATCAAGATACCTGATATCTGGAGGTG 480

QY 1469 TGGATTTTTTCAGTCATAATTTGGGACATATTTCTCGAGAAATGAAACATATCTTCTGTG 1528
DB 481 TGGATTTTTTCAGTCATAATTTGGGACATATTTCTCGAGAAATGAAACATATCTTCTGTG 540

QY 1529 TTCAATGGTGGTGAATATCTCAACTTCTAGTTCACCTGAAACTCTAGTGCAGAGTAC 1588
DB 541 TTCAATGGTGGTGAATATCTCAACTTCTAGTTCACCTGAAACTCTAGTGCAGAGTAC 600

QY 1589 AGCACTGCATCTGCTGTAGCCAGTGCACCTCAGTGCAGCTTCTAAGTTTGGAGAGA 1648
DB 601 AGCACTGCATCTGCTGTAGCCAGTGCACCTCAGTGCAGCTTCTAAGTTTGGAGAGA 660

QY 1649 AAAAAATGCATAATGTTGGCATCTCGTCACCTTTTCTTCAAGTAATCAAAATGAGGC 1708
DB 661 AAAAAATGCATAATGTTGGCATCTCGTCACCTTTTCTTCAAGTAATCAAAATGAGGC 720

QY 1709 CTTCTGATGATTAACCTGTGTGGGTGTTTCAGATGGTTCGTGTACGTCTGCAATAGG 1768
DB 721 CTTCTGATGATTAACCTGTGTGGGTGTTTCAGATGGTTCGTGTACGTCTGCAATAGG 780

QY 1769 ATACTGGTGCATGGATCGTTGTGTGATGGGATACAGAGTGTGATATCTTAAACGCTT 1828
DB 781 ATACTGGTGCATGGATCGTTGTGTGATGGGATACAGAGTGTGATATCTTAAACGCTT 840

QY 1829 GTGTAAGCTGTTCTGCTGCTGTTGATTTCACTTAGTCAATCCAGCAGTCAACCTAAAAC 1888
DB 841 GTGTAAGCTGTTCTGCTGCTGTTGATTTCACTTAGTCAATCCAGCAGTCAACCTAAAAC 900

QY 1889 AAGCTATGACGAGAGTGTGCTGCTTTAAAAAATATGCCCCATCATTAAGCTACAA 1948
DB 901 AAGCTATGACGAGAGTGTGCTGCTTTAAAAAATATGCCCCATCATTAAGCTACAA 960

QY 1949 CCCTTGCACCTAACTCTTGGCTCTGAGGCAATCTGCAAGGGAATTTACCTAAATAT 2008
DB 961 CCCTTGCACCTAACTCTTGGCTCTGAGGCAATCTGCAAGGGAATTTACCTAAATAT 1020

QY	2009	CTCATAACTCCCTGATGGTTCAAGCAATAAAGACAAACCTTAACAGACCCCGGACATACATG	2068	QY	3089	TGCTTCTGGCGAACTGAGAGAAATTGAGCAGCAGGAGGAAGCAATGATGTCCT	3148
DB	1021	CTCATAACTCCCTGATGGTTCAAGCAATAAAGACAAACCTTAACAGACCCCGGACATACATG	1080	DB	2101	TGCTTCTGGCGAACTGAGAGAAATTGAGCAGCAGGAGGAAGCAATGATGTCCT	2160
QY	2069	TGCTATTTCTTGATGTGGAGCGCTGATATTCAACTCTCTGACTGAAGAAGCCTCTAGGC	2128	QY	3149	GGGCTCTTTACTTTCCTCAGTACATAGACCAAGTCAATCACTGAGTGCATCATCAAGAG	3208
DB	1081	TGCTATTTCTTGATGTGGAGCGCTGATATTCAACTCTCTGACTGAAGAAGCCTCTAGGC	1140	DB	2161	GGGCTCTTTACTTTCCTCAGTACATAGACCAAGTCAATCACTGAGTGCATCATCAAGAG	2220
QY	2129	CGAATACTGCTCTTATTTTCCCAAGAGAAATTGCAAAAAGCATCTGGCAGTTTCAGACAAAG	2188	QY	3209	CGCGCAGACTATACACACCGCTCTGATGCTCAGGGCTGAAGCAAAAGTCCAGGAGS	3268
DB	1141	CGAATACTGCTCTTATTTTCCCAAGAGAAATTGCAAAAAGCATCTGGCAGTTTCAGACAAAG	1200	DB	2221	CGCGCAGACTATACACACCGCTCTGATGCTCAGGGCTGAAGCAAAAGTCCAGGAGS	2280
QY	2189	GGGGCTCTTTTAACTGGAAAAAGCAGCAGCTCTCTTCCAAACCAAGTGAAGAAACGA	2248	QY	3269	AAGAGCATCAGCTTGTGTGACGATGACATCACTGCTGCTTATCAAGTGTCCCAAAA	3328
DB	1201	GGGGCTCTTTTAACTGGAAAAAGCAGCAGCTCTCTTCCAAACCAAGTGAAGAAACGA	1260	DB	2281	AAGAGCATCAGCTTGTGTGACGATGACATCACTGCTGCTTATCAAGTGTCCCAAAA	2340
QY	2249	TCAAAGAGACATCAAGGAACCTCTCTGATGATGAAGAGGAGGATGAGGAGATAATGA	2308	QY	3329	TGAAAAAAATTTTCTACATCTTACGAGAAAGACGGAAGCAAGCTACCGCTATTGTTTTAC	3388
DB	1261	TCAAAGAGACATCAAGGAACCTCTCTGATGATGAAGAGGAGGATGAGGAGATAATGA	1320	DB	2341	TGAAAAAAATTTTCTACATCTTACGAGAAAGACGGAAGCAAGCTACCGCTATTGTTTTAC	2400
QY	2309	GGCAGAGAGGGAAGAAAGTGATCTCTGAATATCGGTCCAGCAAAATCAAAGCCATTGACCC	2368	QY	3389	TTGGAGTATAGAGCTGAATTTGGTGCTGAATTTGAACCTCTCTAAACTATTGACAGAC	3448
DB	1321	GGCAGAGAGGGAAGAAAGTGATCTCTGAATATCGGTCCAGCAAAATCAAAGCCATTGACCC	1380	DB	2401	TTGGAGTATAGAGCTGAATTTGGTGCTGAATTTGAACCTCTCTAAACTATTGACAGAC	2460
QY	2369	TATTAGATAATAATTTAACTATGACACTGCAAAAGCTGTTTATGCTGCTGCTTACGCGCT	2428	QY	3449	CTCGAAGCTCTAGCCAAAATTCCTGAGGGATTCGGGTGACTAGTGGTGGATCCAACTACT	3508
DB	1381	TATTAGATAATAATTTAACTATGACACTGCAAAAGCTGTTTATGCTGCTGCTTACGCGCT	1440	DB	2461	CTCGAAGCTCTAGCCAAAATTCCTGAGGGATTCGGGTGACTAGTGGTGGATCCAACTACT	2520
QY	2429	GGGGTTTGAATGAAGTACTGGAATGAAATTTGCTGGATGCTTGAATGCTCAAAACCCC	2488	QY	3509	CGTGGCCAGACATCTTGCAAAGGCACTGACGCTTCTCTGCTACAGCCTCCAAAGCCCCA	3568
DB	1441	GGGGTTTGAATGAAGTACTGGAATGAAATTTGCTGGATGCTTGAATGCTCAAAACCCC	1500	DB	2521	CGTGGCCAGACATCTTGCAAAGGCACTGACGCTTCTCTGCTACAGCCTCCAAAGCCCCA	2580
QY	2489	ACTGCAACCGTATCGTTTGGCTCTTGTCAAGAGGAGGCCATATGTCACCTGATGCTGCCG	2548	QY	3569	AACTTCTCTCACACAGCAGCTATCCGAGAAAGACCCATTCATCTGATGGAAGTGGGTTCA	3628
DB	1501	ACTGCAACCGTATCGTTTGGCTCTTGTCAAGAGGAGGCCATATGTCACCTGATGCTGCCG	1560	DB	2581	AACTTCTCTCACACAGCAGCTATCCGAGAAAGACCCATTCATCTGATGGAAGTGGGTTCA	2640
QY	2549	GTTTAAATCAGCTGCTGTTAACTGTCATCGGAAACAGAAAGTAGGAGAGAGCTGC	2608	QY	3629	CTGTTTGGAGGCTTACATGGATGCTCCGCTGTTCTGATGGGGCTTCTCGAACTTTTGTG	3688
DB	1561	GTTTAAATCAGCTGCTGTTAACTGTCATCGGAAACAGAAAGTAGGAGAGAGCTGC	1620	DB	2641	CTGTTTGGAGGCTTACATGGATGCTCCGCTGTTCTGATGGGGCTTCTCGAACTTTTGTG	2700
QY	2609	CAGGCTGAGGAGTAGGAAAGGAACTTACGGAGTGCCCGTGGCCGTCACACACAGC	2668	QY	3689	CGGATCCGAGAAACAACTTGCACATCAATGAGGGTTCCTCTGAGCCACAGCAGCTG	3748
DB	1621	CAGGCTGAGGAGTAGGAAAGGAACTTACGGAGTGCCCGTGGCCGTCACACACAGC	1680	DB	2701	CGGATCCGAGAAACAACTTGCACATCAATGAGGGTTCCTCTGAGCCACAGCAGCTG	2760
QY	2669	ATCTCTGCTCTATCATTTCTTGGCAAAATCTTTAATGAGTATGACCAATGCAAACTTTTA	2728	QY	3749	ACTGGCCCGCTCTGAGGAGCATGCCCTCTGCTCATTCGCCACCCGACAGCACCAGCT	3808
DB	1681	ATCTCTGCTCTATCATTTCTTGGCAAAATCTTTAATGAGTATGACCAATGCAAACTTTTA	1740	DB	2761	ACTGGCCCGCTCTGAGGAGCATGCCCTCTGCTCATTCGCCACCCGACAGCACCAGCT	2820
QY	2729	TTGGTGATCATATGAGAGGGTCTTACAGGCCACCTAGACCCAGCAGCAGCTTT	2788	QY	3809	TCATCACCACTAGCAGGAGGTACACAGACATACGGCTCTTTCAGGCAAAATACCCAAT	3868
DB	1741	TTGGTGATCATATGAGAGGGTCTTACAGGCCACCTAGACCCAGCAGCAGCTTT	1800	DB	2821	TCATCACCACTAGCAGGAGGTACACAGACATACGGCTCTTTCAGGCAAAATACCCAAT	2880
QY	2789	CTAAGGCAAGGGTCCCTCCAACTTCCAGTAAATTTGTCAGGAGGACAGATTAACAAG	2848	QY	3869	CACAGCAGAAATATGCAACAACAACTCTTGCACAGAGCTTAAAGGGGAAAATTTTCAGAGTCA	3928
DB	1801	CTAAGGCAAGGGTCCCTCCAACTTCCAGTAAATTTGTCAGGAGGACAGATTAACAAG	1860	DB	2881	CACAGCAGAAATATGCAACAACAACTCTTGCACAGAGCTTAAAGGGGAAAATTTTCAGAGTCA	2940
QY	2849	TTGCTGACCTGCTGCTTCCGCTCGGTCTGATGCTGATCACTCTGCGCTTGACCCCTCTT	2908	QY	3929	TTGAAATTTCTTATTTGAAAGATGCCACAGATGTTTGGATCTTCTCTGAGGAGTTATGG	3988
DB	1861	TTGCTGACCTGCTGCTTCCGCTCGGTCTGATGCTGATCACTCTGCGCTTGACCCCTCTT	1920	DB	2941	TTGAAATTTCTTATTTGAAAGATGCCACAGATGTTTGGATCTTCTCTGAGGAGTTATGG	3000
QY	2909	CTGCTCTGCTTATCATACCTGTTTCTTAGTAAATGAAAGTTGAGTCACTAGCTGCTA	2968	QY	3989	ACATCATTTATGTAAGGATCTTTAGTAAAGAAAGAGTCTTTCAGAGATGTT	4048
DB	1921	CTGCTCTGCTTATCATACCTGTTTCTTAGTAAATGAAAGTTGAGTCACTAGCTGCTA	1980	DB	3001	ACATCATTTATGTAAGGATCTTTAGTAAAGAAAGAGTCTTTCAGAGATGTT	3060
QY	2969	TGCACTGCTGTTATGCTCCAGACCTACTGGGATTTGATAAATTTAGGCTTCCCTCTCGG	3028	QY	4049	TCCAGCCATCTGACAGGTTTACATGCTCAGCTTATTTATGAGCGGAATTCACAGAAATAGCAG	4108
DB	1981	TGCACTGCTGTTATGCTCCAGACCTACTGGGATTTGATAAATTTAGGCTTCCCTCTCGG	2040	DB	3061	TCCAGCCATCTGACAGGTTTACATGCTCAGCTTATTTATGAGCGGAATTCACAGAAATAGCAG	3120
QY	3029	AGATGCTGGCCCGAAGATGGCAAGATCGATGCTTTGGAGGTGAGAGAGCCGACAGGCC	3088	QY	4109	TTGAGCTCGCCATGTTTCACTGAGTGGCTTGCAGCATCCGCACTCGGAATGCAAGAAATGTCAGCAA	4168
DB	2041	AGATGCTGGCCCGAAGATGGCAAGATCGATGCTTTGGAGGTGAGAGAGCCGACAGGCC	2100	DB	3121	TTGAGCTCGCCATGTTTCACTGAGTGGCTTGCAGCATCCGCACTCGGAATGCAAGAAATGTCAGCAA	3180
				QY	4169	TCCATGGAACAAGGGACCAATCACTGAGTGGCTTTTGTCTCTGATGGAAGATATCTTG	4228


```
Db 3181 TCCATGGACACAGGACCAATCACTGCAGTGGCTTTTTCCTGATGGAAGATATCTTG 3240
Qy 4229 CCACCTACTCAAACTGACAGCCACATTTCTTTTGGCAGATGAACACGTCAGTCTGG 4288
Db 3241 CCACCTACTCAAACTGACAGCCACATTTCTTTTGGCAGATGAACACGTCAGTCTGG 3300
Qy 4289 GAAGCATCGGATGCTGACCTCGGACCTCAGCTGCGTGCATTAACCTACAGGTGC 4348
Db 3301 GAAGCATCGGATGCTGACCTCGGACCTCAGCTGCGTGCATTAACCTACAGGTGC 3360
Qy 4349 CCCCTGTGACCGCGTCCCGCGCTCCCAATGCCCTCAAGCTGCGCGGCTCATCT 4408
Db 3361 CCCCTGTGACCGCGTCCCGCGCTCCCAATGCCCTCAAGCTGCGCGGCTCATCT 3420
Qy 4409 GGACTTCCAAACCGCAAGTCACTCTCATGCGCCCATGAGCGGAAGGAGCACCGCTTCATGG 4468
Db 3421 GGACTTCCAAACCGCAAGTCACTCTCATGCGCCCATGAGCGGAAGGAGCACCGCTTCATGG 3480
Qy 4469 TCTAA 4473
Db 3481 TCTAA 3485

RESULT 10
AF192379 4371 bp mRNA linear ROD 15-SEP-2001
DEFINITION Rattus norvegicus TGF-beta resistance-associated protein TRAG
ACCESSION AF192379
VERSION AF192379.1 GI:15624076
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE TRAG: a novel gene associated with TGF-beta resistance
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1999) LBC, NIH, 37 Convent Drive, Bethesda, MD
20892, USA
FEATURES
source
1..4371
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/cell_type="epithelial"
/tissue_type="liver"
1..>4371
/gene="Trag"
1..>4371
/gene="Trag"
/codon_start=1
/product="TGF-beta resistance-associated protein TRAG"
/protein_id="AA03984.1"
/db_xref="GI:15624077"
/translations="MAGNSLVLPVLWGRKAPTHCISILLDDGGTIVTGCHDQIC
LWLSLEYNRPALLPFTAAITCLSKACASGDQKQVTSASANGMCLWDVNDGRCL
EFTKLACTHTG10PQPSVGNQREGRLCHGYPEILVVDATSLVLSVSKISPDW
ISSMIIIRSHRTQEDTVALSVTGLKVLVYTSIEISGLDTEPIFEESSKPIYCONCO
SLSCFATQSRLLVVCSSKRWRFADGYSLLCSGPDGCTGTGGFVSADKVLITE
NGQSYIYKLSPASCLPDSFRSDVGRKAENLIPVQHSLLDQKDRLEVICPPVTRFFY
CKEYLLKGLDSSGRLSIWNIADTADKQANEGKLTTCISLQDAFDKPKCPAG
I1DQSLIPNSPFLPKYASVYIPAHGRVLCGRDGSIIIVPATQAIVOLLOQEHMFL
RGWPHRUTLRHKNKVTCLLYPHOVSRVDRVLISSGVDFSVI1WDIFSGMKHIF
CVHGGEITQLLVPPENCARSVQHCVSASDHSVGLLSREKICIMLASRHLPIQVIV
KWRPSDYL1VVGCTDGSVCVWQMDTGALDRCAMGITAVEILINACDEAVPAVDSLSHP
AVNLKQAMTRRS1AALKNMHHKILQTLATNLASBASDSKGNLPKYSHNSLMWQAIKTN
```

ORIGIN

```
Query Match 75.6%; Score 3382.8; DB 10; Length 4371;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 3820; Conservative 2; Mismatches 549; Indels 99; Gaps 1;

Qy 1 ATGCGAGGAAACAGCCTTGTCTACCCATTTGTTTGGGGTGGAAAGCGCCACACAT 60
Db 1 ATGCGAGGAAACAGCCTAGTCTGCCCCATTTGTTCTGCGGCGCGCAAGCACACAC 60

Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGAGGGCCACAGATCGTAAACAGATGTACGAC 120
Db 61 TGTATTTTCGTCATCTGCTGACAGATGATGGGGGACAGATGTGTGACTGGATGCGAC 120

Qy 121 GGACAAATATGCTCTGGGATCTTTCAGTAGAAGTGCATAATTAATCTCCAGGACTGTG 180
Db 121 GGACAGATATGCTCTGGGATCTCTCAGAAGAGTGGAGTTAATCCCGAGCACTATTA 180

Qy 181 TTTGGTCATACAGCATCAATCACTTGTCTCTAAAGCTTGTCTTCCAGTGACAAACAG 240
Db 181 TTTGGTCACACAGCAGCATCACTTGTCTCTCAAGCCTGTGCTCTGGAGACAAACAG 240

Qy 241 TATATTGTGATGATCTGAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTGAGGAGA 300
Db 241 TACACCGTGAGTGGCTCGCAAAATGGAGAGATGTGCTCTGGGATGTGAAATGACGG 300

Qy 301 TGTATTGAATTTACAAAATTAGCTTGCACACATCTGGCATACAGTTCTTACCAGTTCT 360
Db 301 TGTATTGAATTTACAAAATTAGCTTGCACACATCTGGCATACAGTTCTTACCAGTTCT 360

Qy 361 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCCACGACATTAACCTCGAAATCCTTGT 420
Db 361 GTTGGAAATCAGCGAGAGGAGGCTTCTGTGCCATGGGCATTAACCTCGAAATCCTGT 420

Qy 421 GTGGATGCTACACGCTTGAAGTATTATATCTCTTAGTATCAAGATATCAACAGACTG 480
Db 421 GTGGATGCCACACGCTTGAAGTATTATATCTCTTAGTATCGAAGATCTCTCCAGACTG 480

Qy 481 ATTAGCTCCATGATGATTTATTCGATCCCGGACACAGGAGCACAGTGGTAGCACTC 540
Db 481 ATCAGCTCCATGATGATTTATTCGATCCCGGACACAGGAGCACAGTGGTAGCGCTG 540

Qy 541 TCGGTGACTGGCATCTCGAAGGTCTGGATTGTGTACCTCGGAAATAAGTGACATCGAGAT 600
Db 541 TCGGTGACAGGTATTCTGAAGGTCTGGATTGTGACCTCTGAAATTAGTGGATGCGAG 600

Qy 601 ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTTGTGAGATTCGCCAAAGCAT 660
Db 601 ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTTGTGAGATTCGCCAAAGCAT 660

Qy 661 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTTCCTTCCAAATATTTCGAGGGTG 720
Db 661 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTTCCTTCCAAATATTTCGAGGGTG 720

Qy 721 TTGATGCTGGGAGACTTATCTTGTGTGTGTTCAGGTCTTAGTGAATAGCAGACATGG 780
Db 721 TTGATGCTGGGAGACTTATCTTGTGTGTGTTCAGGTCTTAGTGAATAGCAGACATGG 780

Qy 781 ACCGGGGGGGACTTTTGTCTCATCAGATAAAGTCATCATTTTGGACAGAAAATGGGCAAGT 840
```

```
LTDPDIHVLFFDEVALIIQLTTEASRPNTALLSPENLQKASGSSDKGGSFLTKGRRAA
VLFQVKETIKENIKHLEDEEBEEMRREESDPYRASKSPKLTLLLEYNLATWD
TAKLFLMSEKLENGLEVDCLDRGLMKLPHCTVSPFLSRGHHMSLMLPLFVNNQAG
KLQAKAEARAGKGPATSVGKTVTVSRVATTOHLLSIISLANHMGMTNATFTGDHM
KKGPTPRPGTPTDLKARDSPPASSNIVQOIKQWSQLAAHVCNMLPDLGLGAKPR
PILLEMARWODRCLEVEEAQALLAELERRLEQAGKETIDTAAPYLQPYMDHVLIS
PGVTAEMQTMAAAAPDASGPEAKVQEEHLEVDITTTGCLSSVPQMKKSTSEEBR
KQATAVLLGVIGABFABETPPKLLTRPRSSQIPGFLTGGSSNYSLARHTCKAL
TFLLOPPSPKLPPTSHRTTADLIGRGTVMPEYDVSALVHMLGLCELCADEKQAL
NIWGLPLSPADSARSARHSLIATARPAPTITTAKEVHRHTALAAANTQSOQSH
TTTLARAKGBILRVETILIEKMPDVLVLADEVMDIIMYCLEGLSVKKGKLOECFPAI
CRFVMSYVESHRIVAGBSGSVALYDIETGKQTHGHKGPITATVSPADRYLAT
YSNTDHSISFWQNTSLSGIMNSAPQLRCIKTYQVPPVPASPASPNALRLARLI
WTSNRNVILMAHDGKHBFMV"
```

Db	781	ACTGGAGGACCTTTGCTGTCAGACAAAGTCAATATTTGGACTGAAACGGCGCAGAGT	840	Db	1861	CTCAGTCACCCAGCAGCTCAACCTGAGCAAGCCATGACACGGCGGAGTCTCGCGCCCTT	1920
Qy	841	TATATTACAACTACCTGCCAGTGGCTTCCAGCTAGTGATTCATTCGCGAGTGATGG	900	Qy	1921	AAAAATATGGCCCATCATTAAGCTACAAACCCCTTGCACTAAACCTCTTGGCTTCTGAGGCA	1980
Db	841	TACATTTACAACTCCCTGCCAGTTGGCTTCCAGCTAGTGATTCATTCGCGAGTGACGTG	900	Db	1921	AAAAATATGGCCCATCACCAAGCTGCAACCCCTTGGCACTAAACCTTTTGGCTTCTGAGGCC	1980
Qy	901	GGGAGGCAGTTGAAATTTAATTCCTCTGTTACACATATCTCTTGGATCGAAAGAT	960	Qy	1981	TCTGCAAGGGGAAATTTTACCTAAATATCTCTCAATACTCCCTGATGCTTCAAGCAATAAAG	2040
Db	901	GGGAAAGCAGTTGAAATCTGATTCCTCCGCTGCGCATAGCCTCTTGGATCGAAGGAT	960	Db	1981	TCTGCAAGGGGAAATTTTCTAAATATCTCTCAATACTCCCTGATGCTTCAAGCAATAAAG	2040
Qy	961	AAAGAGTTGCTAATTTGCTCTCTGTTACTCGGTTCTTATGATGACGAGAAATATTC	1020	Qy	2041	ACAAACCTTAACAGACCCGAGCATACATGTGCTATCTTTTGAATGTGGAAGCTTGATAT	2100
Db	961	AGAGAGTTGCTAATTTGCTCTCTGTTACTCGGTTCTTATGATGACGAGAAATATTC	1020	Db	2041	ACAAACCTTAACCTGACCCGGATATCCATGTGCTTTCTTTGATGTGGAAGCTTTGATAT	2100
Qy	1021	CATAAACTGTTAATTCAGGGTGATTTCTTCTGGAAGTTGAATATTTGGAAATATCAGAC	1080	Qy	2101	CAACTCTGACTGAAAGAGCCTCTAGCCGGAATACCTCTTATTTTCCCGAGAGATTCG	2160
Db	1021	CATAAGCTACTAATTCAGGGTGATTTCTTCTGGAAGTTGAATATTTGGAACATAGCAGAC	1080	Db	2101	CAACTCTGCTGAAAGAGCCTCTAGCCGGAATACCTCTTATTTTCCCGAGAGATTCG	2160
Qy	1081	ACAGCTGATTAACAGGGAAGTGAAGAGGGCTGGCAATGACAACTTCTATTTAGTTTGC	1140	Qy	2161	CAAAAAAGCATCTGGCAGTTTTCAGAAAGGGGGCTCTTTTAACTTGGAAAAACGAGCAGCA	2220
Db	1081	ATAGCAGACAAACAGGAAGCCAAATGAAGGGCTAAAAACGAACTTGTATTTAGTTTGC	1140	Db	2161	CAGAAAGCATCTGGCAGTTTTCAGAAAGGGGGCTCTTTTCTGACTGGAAAAACGAGCGGCA	2220
Qy	1141	GAGGCAATTCATAACTGAATCTTGTCTGCTGGAATATAGATCAGCTGAGTGTGAT	1200	Qy	2221	GTTCCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACCTCTCTTGAT	2280
Db	1141	GATGCAATTCAGCAACTGAAGCCTTGTCTGCTGGAATATAGATCAGCTGAGTGTGAT	1200	Db	2221	GTTCCTCTTCCAGCAAGTGAAGAAACCTATCAAGAGAAACATCAAGGAGCAGCTCTTGAT	2280
Qy	1201	CCCAATAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTACATACAGACATGAGACGA	1260	Qy	2281	GATGAAGAGGAGATGAGGAGATTAATGAGCAGAGAGAGGAGAGAAAGTGCATCTGATAT	2340
Db	1201	CCAAACAGCAACGAACCACTTAAAGTAACTGCGAGTGTCTACATACAGCAGACCGGGCGC	1260	Db	2281	GAGGAGGAGGAGCAGAGAGAGGATGAGCAGAGAGGAGGAGAAAGTGCATCTGAGTAC	2340
Qy	1261	CTTGTTTGGTCTGGAAGATGGAAGCATAGTATTTGTACTCTGCCACACAGACGGCCATA	1320	Qy	2341	CGGTCCAGCAAAATCAAAAGCATTGACCTATTAGAATAATAATTTAACTATGACACTGC	2400
Db	1261	CTTGTTTGGTCTGGAAGATGGAAGCATAGTATTTGTACTCTGCCACACAGACGGCCATA	1320	Db	2341	CGGGCCAGCAAGTCTCAACCACTTACCTACTAGATACACCTTACTATGAGTACCGCA	2400
Qy	1321	GTACAGCTGTTGCAAGGGGAAACATGCTCAGAGAGGTTGGCCACCTCACAGAACTC	1380	Qy	2401	AAGCTGTTATGCTCTGCTTCAAGCTGCGGGTTTGAATGAAGTACTGGAATGAAGTTTGC	2460
Db	1321	GTTCAGCTGCTGAGGAGAAACATGCTCAGAGAGGTTGGCCCCCGCACAGAACCTC	1380	Db	2401	AAATATTATCATGCTCTCTCCAGCGCTGGGGTTTGAATGAAGTCTGGAATGAAGTTTGC	2460
Qy	1381	CGTGGTCATCGGAACAAAGTCATGTTTGTCTATCTCATCAGGTTCTCAGCTCGGTAT	1440	Qy	2461	CTGGATGCTTGGATGCTGAAACCCCACTGACCGCTACCTGCTTGGCTCTTGTCAAGA	2520
Db	1381	CGTGGCCACCGGAACAAAGTCATGTTTGTCTATCTCATCAGGTTCTCAGCTCGGTAT	1440	Db	2461	CTCGATGCGCTCGCATGCTGAAACCACTGACACAGCTGACAGTGTCTTGTCTCTTATCAGA	2520
Qy	1441	GATCAAGATACCTGATATCTGAGGTTGGATTTTTCAGTCATTAATTTGGGACATATTT	1500	Qy	2521	GGAGGCGCATATGTCATGATGCTGCGGGTTTAAATCAGCGCTCTTGTAACTGTGCACAT	2580
Db	1441	GATCAAGATACCTGATATCTGAGGTTGGATTTTTCAGTCATTAATTTGGGACATATTT	1500	Db	2521	GGAGGTCATATGCTCTTGTGCTTCTCTGTTTAAATCAGGCTCTGGAAGTACTGACAG	2580
Qy	1501	TCTGGAGAAATGAACATATCTCTGTTTCTGTTGATGAGTACTCAACTTCTAGTT	1560	Qy	2581	GGAAAAACAGAAATGAGGAAGAGCTGCCAGCTCTGAGGGAGTAGGAAGGAACTTAC	2640
Db	1501	TCCGGAGAAATGAACATATCTCTGTTTCTGTTTCTGTTGATGAGTACTCAACTTCTAGTT	1560	Db	2581	GCGAAAGCAGAAACGAGACGGAAGGGGCCAGCAACCGAGAGCGTAGGCAAGGGGACCTAC	2640
Qy	1561	CCACCTGAAAATCTAGTGAAGAGTACAGACATCTGCTCTGTAGCCAGTGACAC	1620	Qy	2641	GGAGTGTCCGTCGCTGCCACCAACAGCATCTCTGCTATCATTTCTTTTGGCAATACT	2700
Db	1561	CCGCCAGAAACTCTAGTGAAGAGTTCAACTGCTGCTTCTGTTGCGCAGTGACAC	1620	Db	2641	ACAGTGTCCGTCGCGCTGCTGCAACCTCTGTTGCTCAATCATATCTCTGGCAATACT	2700
Qy	1621	TCAGTAGACTTCTAAGTTTGGAGAGAAAAATGCAATATGTTGGCATCTCGTACCTT	1680	Qy	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGGTCTTACCAGG	2760
Db	1621	TCTGTAGGCTGCTAAGTCTGCGAGAGAAAAATGCAATATGTTGGCGTCTCTCGTACCTG	1680	Db	2701	TTAATGAGCATGACCAATGCGAGTTTCAATTTGAGATCATATGAAGAGGGGCCACAGG	2760
Qy	1681	TTTCCTATTCAAGTAATCAAAATGAGGCTTCTCATGATTAATCTGTTGGTGGGGTGTCA	1740	Qy	2761	CCACCTAGACCAAGCAGACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT	2820
Db	1681	TTCCCTATTCAAGTATCAAGTGAAGGCTTCTGACGACTACTTGTGGTGGGGTGCAG	1740	Db	2761	CCGCTAGACCAAGCAGACCCAGACCTTTCTAAGGCAAGGGATTTCCCGCCAGCCTCCAGT	2820
Qy	1741	GATGGTTCTGTGATCGTCTGCAAAATGATATCTGGTGCATTTGGATCGTTGTGATGGG	1800	Qy	2821	AAATTTGTGCAAGCAGATTTAAACAGTTGCTGCACCTGCTGCTTCCGCTCGTCTGAT	2880
Db	1741	GACGGCTCTGTGTGTCTGGCAGATGGAACACTGGTGGCTGGACCGCTGTGCAATGGG	1800	Db	2821	AAATTTGTGCAAGCAGATTTAAAC	2845
Qy	1801	ATACAGCAGTTGAGATTTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTTGATTC	1860	Qy	2881	GCTGATCACTCTGGCTCTGACCTCTCTTCTGCTCTGCTTTTACATACCTGTTTCTAGTA	2940
Db	1801	ATACAGCCGGGAGATTTCTCAATGCTGTGACGAAGCTGTCCCTGAGCAGTGGACTCA	1860	Db	2846	-----	2845
Qy	1861	CTTAGTCTATCCAGCAGTCAACCTTAAACAAAGCTATGACGAGACCTAGTCTTGTGCTCTT	1920	Qy	2941	AAATGAAGTTGGAGTCAAGTTAGTCTGATGCACTGTGTTATGTGCGCAGACCTACTCGGA	3000
				Db	2846	----AAGGATGGAGCCAACTAGCTGCCATGCACTGTGTCTGCTGCGGAGCCTGCTGGG	2901

```
QY 3001 TTGATTAATTTAGGCTCCCTCTTCTGAGATGCTGGCCCGAAGATGGCATGATCGATGC 3060
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2902 CTGGGTAAATTCAGGCTCTCTTCTGAGATGCTAGCTCGAAGATGGCATCGGTGC 2961
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3061 TTGAGGTGAGAGAGCGGACAGCGCTCTGCTTCTGGCGAATCTGAGAGATTTGACGAG 3120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2962 TTGAGGTGAGAGAGGTGACAGCGCTCTTCTTAGCAGAGCTGAGAGATTTGACGAG 3021
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3121 GCAGCGAGGAGGAGGACATTTGATGCTGGGCTCTTCTTACCTCAGTACATAGACAC 3180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3022 GCAGCGAGGAGGAGATTTTGTATACCTGGGCTCTTCTTACCTCAGTACATAGACAC 3081
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3181 GTCATATCACTGAGTACATCAGAGCGCGGAGCATATACACCGGCTCTGATGCC 3240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3082 GTCATATCACCTGAGTACAGCGGAGGACCATGAGCATATGGCAGCTGCTCCAGATGCC 3141
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3241 TCAGGCGCTGAAGCAAGTCCAGAGGAGAGAGCATGACCTTGTGACGATGACATCAC 3300
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3142 TCGGGGCCAGAGCCAAAGTCCAGAGAGAGAGCATGACCTCGTGGACGATGACATCAC 3201
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3301 ACTGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3360
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3202 ACTGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3261
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3361 CGAAGCAAGTACCGCTATTTGTTTACTTGGAGTAATAGAGCTGAATTTGGTGTGAA 3420
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3262 AGAAGCAGGCACTGCTATTTGTTCTCTGGAGTATAGAGCAGAGTTTGGAGCTGAA 3321
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3421 ATTGAACCTCTTAACCTATGACAGAGCTCGAGCTCTAGCCAAATTCCTGAGGATTC 3480
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3322 ATTGAACCAACCAAACTGCTGACAGAGCTCGGAGCTCTAGTCAAATTCCTGAGGATTT 3381
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3481 GGGTTGACTAGTGTGATCAACTACTCTGCTGGCCAGACATACTTCAAGGCACTGACG 3540
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3382 GGTTGACAGTGGAGTTCCAACTACTCTCTGCGCAGACATAGCTGCAAGGCACTGACA 3441
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3541 TTTCTTCTGTACAGCTTCAAGCCCAAACTTCTCTCCACACAGCATATCCGAAGAACA 3600
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3442 TTTCTTCTGTACAGCCACCAAGTCCCAAACTTCTCTCTCATAGCACCATCCGAGAACT 3501
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3601 GCCATTGATCTGATGAGCTGGGTTCACTGTTTGGAGCTTACATGGATGTCCGCT 3660
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3502 GCCATTGATCTGATGAGCTGGGTTCACTGTTTGGAGCTTACATGGAGTGTCTGCT 3561
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3661 GTTCTGATGGGCTTCTCGAACTTGTGCGGATGCCGAGAAACAACTTGCAACATCACA 3720
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3562 GTCTGATGGGCTGTGGAGCTGTGTGAGATGCTGAGAGCGATGGGCCAATCACA 3621
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3721 ATGGGGTTCCTGAGCCAGCAGCTGACTCGGCGGCTCTGCGAGGATGCCCTCTCG 3780
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3622 ATGGGGTTCCTGAGCCCTGACGCTGACTCTGCGCGATCCGCAAGACACGCCCTTCT 3681
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3781 CTCATTGCCACCGCAGACACCGGCTTCTATCAGCAGCATAGCCAAAGAGGTACACAGA 3840
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3682 CTATAGCCACCGCAGACACCGGCTTCTATCACCACCATAGTAAAGAGGTGACAGA 3741
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3841 CATACGCTCTTGCAAGAAATACCAATACAGCAGAGATATGACAAACAACTCTTGCA 3900
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3742 CACAGCGCTTTCAGCAAAATACCAATCCAGCAGAGATATCCACCAACCACTGGCA 3801
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3901 CGAGCTAAAGGGGAAATTTTGAGAGTCAATTTGAAATCTTATTTGAAAAGATGCCACAGAT 3960
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3802 AGGCTAAAGCGGAAATCCTGAGAGTCAATTTGAAATCTTATTCGAAAAGATGCCACGAT 3861
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3961 GTTGTGATCTTCTGAGGAGGTATGGAATCATATTTAGTACTGCTTGAAGATCTTTA 4020
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3862 GTTGTGATCTTCTTGTGAGGTTCATGGACATCATCATGTACTGCTGGAAGGATCTTTA 3921
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4021 GTTAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCAAGTCTTACATGTCAGC 4080
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3922 GTTAAAGAAAGGCTCTTCAAGAGTGTTCACAGGATGTTTCCAGCCATCTGCAGGTTCTACATGGTCAGC 3981
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 4081 TATTATGAGCGGAATCAGAGAAATGAGTTGGAGCTGCCATGGTTTCAGTGGCCCTGTAC 4140
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3982 TATTATGAGCGGAGTCAAGAAATGCAATTTGAGTTGGAGCAGCCATGGCTCAGTGGCCCTGTAT 4041
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4141 GACATCCGGAATCTGAAAATGTCAGACAATCCATGAGCACAAGGACCAATCCTCAGTGCAGTG 4200
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4042 GACATCCGGAATCTGAAAATGTCAGACATCCAGGACACAGGACCTATCCTCAGTGCAGTG 4101
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4201 GCTTTTGTCTCTGATGGAAGATATCTTGGCCACCTACTCAAACACTGACAGCCACATTTCT 4260
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4102 TCCTTTGTCTCTGATGCGGCTTACCTTTGCCACCTACTCAAACACTGACAGCCACATTTCT 4161
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4261 TTTTGGCAGATGAACAGCTCACTGCTGGGAGGATCGGCACTGCTGAATCGGCACTCAG 4320
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4162 TTTTGGCAGATGAACACCTCACTTCTGGGAGGATTTGGCACTGCTGAATCAGCACTCAG 4221
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4321 CTGCGCTGCATTAACACCTACACAGGTGCCCTCTGTGCAGCGCCGCTCCCGGCTCCAC 4380
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4222 CTGCGCTGCATTAACAGCTTACACAGGTACTTCCAGTGACGCCGCTCCCGGCTCCGAC 4281
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4381 AATGCCCTCAAGCTGCGCCGCTCATCTGGAATTTCCAAACCGCAACGCTCATCTCATGGCC 4440
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4282 AACGCCCTCAGTTGGCCGCTCACTGGAATTTCCAAACCGCAATGTTATCTCATGGCC 4341
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4441 CATGACGGGAGGAGCAGCCTTTCATGGTC 4470
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4342 CACGATGGGAGGAGCAGCCTTTCATGGTC 4371
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AF188123 4371 bp mRNA linear ROD 15-SEP-2001
LOCUS Mus musculus TGF-beta resistance-associated protein TRAG (Trag)
DEFINITION mRNA, partial cds.
ACCESSION AF188123
VERSION AF188123.1 GI:15624070
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4371)
AUTHORS Sanders,S. and Thorgeirsson,S.S.
TITLE TRAG: a novel gene associated with TGF-beta resistance
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4371)
AUTHORS Sanders,S. and Thorgeirsson,S.S.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1999) LEC, NIH, 37 Convent Drive, Bldg. 37, Rm
3C28, Bethesda, MD 20892-4255, USA
FEATURES
Location/Qualifiers
1..4371
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129svj"
/db_xref="taxon:10090"
/chromosome="18"
/map="18D.1-R.3"
/tissue_type="liver"
1..54371
/gene="Trag"
1..54371
/gene="Trag"
/codon_start=1
/product="TGF-beta resistance-associated protein TRAG"
/protein_id="AA03981.1"
/db_xref="GI:15624071"
LWDVSVLBNPRAILLFGHTASITCLSKACAGDKRYTVDSAGANGMCLWDVNDGRCI
BPTKLACTHTQFYQFVSGNOQEGRLCHGHYBELLVVDATSLVLSVSKISPDW
ISMSIHSQRTQEDTVALSVTLKWIIVTSEMSGMQDTPIFREESKPIYQNCQ
SISFCAFTQSLVWCKYRVDFDAGDYSLLCGSPSENGQWTGDFGVFSADKVIWTE
NQGSYTYKLPASCLPASDSFRSDVGKAVENLIPPVQHSLLDQKDELVI CPPVTRFFY
```

GCKEYLHLLKIQDSSRLNWNIAIEKQREADGLKWTTCISLQEAFFDKLKPAPAG
IIDLQSVIPNSNEPLKVTASVYIPAHGRYVCGREDGSIILVPAQTATVJVLQGEHML
RRWPHRTLLGPHNRNKCITCLLYPHQVSARDYDQRYLLISGGVDFSVIWDLPSEMKHIF
CVHGSEITQLLLVHNRNKCITCLLYPHQVSARDYDQRYLLISGGVDFSVIWDLPSEMKHIF
KWRSDYLVVGSSTDSVYVMDTGALDRGAMGITAIVELNACDRAVPAVADLSHP
AVNLKQAMTFRSLAALKNNWAHKLQTLATNLLASEASDKGNLPKYSHNSLMVQAIKTN
LTPDQTHLFPDENKEHLIOLLETSRPNATLISPENLOKAGSSDKGGSFLTGRKAA
LVPOQVKETIKENIHEHLLDEEEDREARQRESDPEYRASKSPKPLTLLVNLNMD
TAKLFMSCLHWGLNEVLDEVLCDRLGMLKPHCTVSGFLLSRGSHMSLMLPGYNQAG
KLHAKAEVGRKLPAAEGVGKGYTVSRAVTTTOHLUSIISLANTLMSHLLPFIQDHR
KKGPTPRPGRKLPAAEGVGKGYTVSRAVTTTOHLUSIISLANTLMSHLLPFIQDHR
PPLLEAMRQWDRLEVEAQAALLAEHLRIEQAQREKIDTWAPLYLPQYMDSHVIS
PGVTAEAMTAAADPASPEAKVQBEEDHLDVDDITAGCLASVPOMKKISYEER
KQATAYLVILGIVGAEGABIEPKLLTRPRSSQIPEGGLTSGGSSNYSARHTCKAL
TYLLQPPSPKLPSPHSTIRKATDLIGRGFTVWEPYMDVSAVIMGLLELCADEKQOLA
NITMGLPLSPADSRASRHALSIATAPFPAPITIAKEVHRHTALAAATQSQSIH
TTTLARAKGEILVBIILIEKMPDQVLLVEVMDIIMYCLEGSLVKKGLQAECPFAI
CRFVMSYVYERSHRIAVGARHGSVALYDIRTKCQTHGHKGPITAVSPAPDRYLAT
YSNTDHSIISFWMQNTSLGSIWMLNSAPQLRCIKTVQVPPVPASPQSHNALKLARLI
WTSNRNVLMHMDKGEHRPMV"

ORIGIN

Query Match	75.3%;	Score 3369.2;	DB 10;	Length 4371;
Best Local Similarity	85.3%;	Pred. No. 0;		
Matches 3813;	Conservative 0;	Mismatches 558;	Indels 99;	Gaps 1;
Qy	1	ATGCGAGGAAACAGCCTTGTTCTACCCATGTTCTTTGGGGTGCAGAAAGGCCACACAT	60	
Db	1	ATGCGAGGAAACAGCCTAGTGTCTGCCCATGTTCTTTGGGGCGGCAAGACCCACACAT	60	
Qy	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTCCAGC	120	
Db	61	TGCATTTCTGTCATCTGTTGACAGATGATGGGGGCACAATTTGTAATCTGGATGCCACGAT	120	
Qy	121	GGACAAATATGTCTCTGGGATCTTTTCACTAGTAACTGCAAAATTAATCTCTCGAGACCTGTTG	180	
Db	121	GGACAAATATGTCTCTGGGATGTTTTCGGTAGAACTAGAAATTAATCCCCGAGACCTGTTA	180	
Qy	181	TTTGGTTCATACAGCATCAATCACTGTTTGTCTTAAAGCTTGTCCTCCAGTGAACAAACAG	240	
Db	181	TTTGGCCACACAGCATCCATCACTGTTTGTCTAAAGCCTCGCCTTTCTGGGGCAAGCGG	240	
Qy	241	TATATTGTGAGTGCATCTGAAGTGGAGAGATGTGCTCTCGGATGTGAGTGAATGGCAGA	300	
Db	241	TAACTGTGAGCGGTCTGCAACCGAGAGATGTGCTCTGGGATGTGAACGATGGCAGA	300	
Qy	301	TGTATTGAATTTACAAATTTAGCTTGCCATACATCTGGCATAAGTTCTACAGTTCTCT	360	
Db	301	TGTATTGAATTTACCAAGTTAGCTTGCCACATCACTGGCATAAGTTCTACCAAGTTCTCT	360	
Qy	361	GTTGGGATCAGCGAGAGGAGGCTTTTATGCCACGACATTAACCTGGAATCCTTGTT	420	
Db	361	GTTGGGAAACAGCAAGAGGCGAGGCTCCTCTGCCATGACATTAACCTGGAATCCTCGTT	420	
Qy	421	GTGATGCTACAGCCTTGAAGTATTTATCTCTTATGATCAAGATATCAACAGACTGG	480	
Db	421	GTGATGCCACAGCCTTGAGTGTGTATTTCTTGTGATCGAAGATCTCTCCAGACTGG	480	
Qy	481	ATTAGCTCCATGAGTATTTATTCGATCCACCGAACACAAGAGGACACAGTGGTAGCATC	540	
Db	481	ATCAGCTCCATGAGCATCATCTCTCAGCGGACACAAGAGGACACTGTGGTGGCGCTC	540	
Qy	541	TCGGTGAATCAGCGAGAGGCTTTTATGCGACGACATTAACCTGGAATTAAGTACATGAGGAT	600	
Db	541	TCTGTGACAGGATTTCTGAAGGTGTGATTTGTCACTCTGGAATTAAGTGAATGCAAGGAT	600	
Qy	601	ACTGAGCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATTCGCAAGCATC	660	
Db	601	ACTGAGCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATTCGCAAGCATC	660	
Qy	661	TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTCCAAATATTTGAGGGGTG	720	
Db	661	TCTTTTGTGCAATTTACACAGAGGTCACTTTTGGTGTGTGTTCCAAATATTCGAGGGGTG	720	

Qy	721	TTTCGATGCCGAGACTATTTCTTTGTTGTTTCAGGTCTTAGTGAAATGACAGACATGG	780
Db	721	TTTGTGCTGGGACACTCTCTCTGTTGCTCAGGTCTTAGTGAAATGACAGACATGG	780
Qy	781	ACCGGGGGGAGACTTTTGCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT	840
Db	781	ACTGGAGGGGACTTTTGCTCTCAGACAAAGTCATCATCTGGACAGAAAATGGGCAAGT	840
Qy	841	TATATTCAAACTACCTGCCAGTTGCCCTTCAGCTAGTGTATTCCTCCGAGTGTATG	900
Db	841	TACATCTACAACTACCTCCAGTTGCCCTTCAGCTAGTGTATTCATTTCCGCGAGCGCTG	900
Qy	901	GGGAGGCGAGTTGAAAATTTAATCTCTCTGTGTAACATATCTCTTTGGATCGAAAAGAT	960
Db	901	GGGAAAGCAGTGGAAAATTTGATCCCTCTCTGTGAGCATAGCTCTTTGGATCAGAAAAGAT	960
Qy	961	AAAGAGTTCTAAATTTTGTCTCTCTGTACTCGGTTCTCTTATGGATGACAGAAATTTTC	1020
Db	961	AAAGAGTTGGTAATTTGTCTCTCTGTACTCGGTTCTCTA CGGATGCAAGGAATATTTG	1020
Qy	1021	CATAAACTCTTAAATTCAGGGTGATTTCTCTGGAAGGTTGAAATATTTTGGAAACATATCAGAC	1080
Db	1021	CATAAGCTACTAAATTCAGGGCGATTTCTCTTGGAAAGGTTAAATATCTGGAACATAGCAGAC	1080
Qy	1081	ACAGCTGATAAAACAGGAAAGTGAAGAAGGCTGGCAATGCACAACTTCTTATTTAGTTGCAA	1140
Db	1081	ATAGCAGAGAAACAGAGACCGATGAAGGCTTAAGATGACAACTTGTATTTAGTTGCAA	1140
Qy	1141	GAGGCAATTTGATAAACTGAATCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1200
Db	1141	GAGGCAATTTGACAGCTGAAGCTTGCCCTCTGCTGGAATTTATCGATCAGCTAAGTGTGAT	1200
Qy	1201	CCCAATGATTAAGAACTCTTAAAGTAACTGCAAGTGTGTGTA CMTACAGCACAATGACGA	1260
Db	1201	CCCAACAGCAATGAACCTCTTAAAGTAACTGCGAGTGTCTACATACAGCACA CCGGCGA	1260
Qy	1261	CTTGTGTTGGTCTGTAAGATGMAAGCATAGTTATTTGTACCTGCGCACACAGCGGCATA	1320
Db	1261	CTTGTGTTGGCGCGGAGATGGAAGCATATTTATGTCCTGCGCACCCAGCGGCATA	1320
Qy	1321	GTACAGCTGTGTCAGAGGGGAAACACATGCTCAGAAAGGTTGGCCACTCACAGAACACTC	1380
Db	1321	GTTCAGCTATTGCAAGGAGAACACATGCTCAGAAAGGTTGGCCCTCACAGAACCTCTC	1380
Qy	1381	CGTGTCTATCGGAAACAAAGTCATCTTTGCTATATCTCTCATCAGCTCAGCTCCGAT	1440
Db	1381	CGTGGCCATCGGAAACAAAGTCACGTGCTGCTGATATCCCATCAGGTCTCAGCTCCGAT	1440
Qy	1441	GATCAAGATATCTGATATCTGGAGGTGTGGAATTTTTCAGTCATATTTTGGGACATATTT	1500
Db	1441	GACCAAGATATCTGATATCTGGAGGTGTGGAATTTTTCGTCATATTTTGGGACATATTT	1500
Qy	1501	TCGAGGAAATGAAACATATCTTCTGTGTCATGTTGGTGTGAGATTA CTAACCTTCAGTT	1560
Db	1501	TCGAGGAAATGAAACATATCTTCTGTGTCATGTTGGTGTGAGATTA CTAACCTTCAGTT	1560
Qy	1561	CCACTGAAAACCTGTAGTGAAGATTA CAGCACTGCTCTGTCTGTAGCCAGTGAACAC	1620
Db	1561	CCCCCGGAAAACCTGTAGTGAAGATTA CAGCACTGCTCTGTCTGTAGCCAGTGAACAC	1620
Qy	1621	TCAGTAGGACTTCTTAAGTTTGGAGAGAAAATGATTAATGTTTGGCATCTCCGTCACTTT	1680
Db	1621	TCTGTAGGCTCTCTAAGTCTGAGAGAGAAAATGATTAATGATTAATGTTTGGGCTCTCCACCTC	1680
Qy	1681	TTTCTTATTTCAAGTAATCAAAATGGAGCCTTCTGATGATTA CTTGTTGGTGTGGTGTTC	1740
Db	1681	TTTTCTTATTTCAAGTCAAGTGGAGCCTTTCGAGCAGCTACCTCGTGTGGTGGATCCACA	1740
Qy	1741	GATGGTCTGTGTACCTCTGGGAAATGGATACCTGGTGCATTTGGTGTGTTGTGATGGGG	1800
Db	1741	GACGGCTCTGTGTATGCTCTGGCAGATGGACACTGGTGGCTGGATGCTTGTGTCAATGGGA	1800


```
Db 3862 GTGGTGGATCTCTTGTGGAGGTGATGATATCATCATGTAATGCTGCTGGAAGGATCTTTA 3921
Qy 4021 GTTAAAGAAAGGTCTTCAAGAAATGTTTCCAGCCATCTGCGAGTTCTACATGTCAGC 4080
Db 3922 GTGAAAAAGAGGCTCTTCAGGAGTGTGTTTCCAGCTATCTGCAGGTTCTACATGTCAGC 3981
Qy 4081 TATTATGAGCGGAATCACAGAATAGCAGTGTGGAGCTCGCCATGTTTCAGTGGCCCTGTAC 4140
Db 3982 TATTATGAGCGCAGTACAGAAATCGCAGTTGGAGCAGCCATGGCTCAGTAGCCCTGTAT 4041
Qy 4141 GACATCGGACTCGAAAAATCTCAGACAATCCATGAGACACAAGGAGCAATCACTGCAGTG 4200
Db 4042 GACATCGGACTCGAAAAATCTCAGACAATCCATGAGCACAAGGAGCAATCACTGCAGTG 4101
Qy 4201 GCTTTTGCTCCTGATGAAGATATCTTGCACCTACTCMAACACTGACAGCCACATTTCT 4260
Db 4102 TCCTTTGCTCCTGATGGCCGTTACCTTGGCCACCTACTCGAACAACCGACGCCACATTTCC 4161
Qy 4261 TTTTGGCAGATGAACAGTCACTGCTGGGAAGCATCGGCATGCTGAACCTGGCCACCTCAG 4320
Db 4162 TTTTGGCAGATGAACAGTCACTGCTGGGAAGCATTTGAAATGCTGAACTCAGCTCCTCAG 4221
Qy 4321 CTGCGCTGCATTAACACCTACACAGTGCCTCCCTGTGAGCGCCGCTCCCGCTCCAC 4380
Db 4222 CTGCGCTGCATTAACACCTACACAGTGCCTCCCTGAGTGCAGCCAGCATCCCTGCTCCAC 4281
Qy 4381 AATGCGCTCAAGCTGGCCCGGCTCATCTGAGCTTCCAAACCGAAACGTCATCTCATGGCC 4440
Db 4282 AAGCGCTTAAGTTGGCCCGGCTCATCTGAGCTTCCAAACCGAAATGTTATCTCATGGCC 4341
Qy 4441 CATGACGGGAAGGACCGCTTCATGGTC 4470
Db 4342 CATGATGGGAAGGAACCGCTTCATGGTC 4371

RESULT 12
AX714633
LOCUS AX714633 3316 bp DNA linear PAT 15-APR-2003
DEFINITION Sequence 1317 from Patent EP1293569.
ACCESSION AX714633
VERSION AX714633.1 GI:29889586
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 1317 19-MAR-2003;
Helix Research Institute (JP); Research Association for Biotechnology (JP)
FEATURES
source location/Qualifiers
1..3316
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 62.8%; Score 2807.4; DB 6; Length 3316;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 3080; Conservative 0; Mismatches 1; Indels 242; Gaps 3;

Qy 1 ATGCGAGGAACAGCCCTGTTCTACCCATTGTTCTTTGGGTCGAAAGCGCCACACAT 60
Db 236 ATGCGAGGAACAGCCCTGTTCTACCCATTGTTCTTTGGGTCGAAAGCGCCACACAT 295
Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGGTAAACAGGATGTCAGAC 120
Db 296 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGGTAAACAGGATGTCAGAC 355
```

```
Qy 121 GGCAAAATATGTCCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCTCTCGAGCACTGTTG 180
Db 356 GGCAAAATATGTCCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCTCTCGAGCACTGTTG 415
Qy 181 TTTGGTTCATACAGCATCAATCACTTGTGTTGTCTAAAGCTTGTCTCCAGTGCAAAACAG 240
Db 416 TTTGGTTCATACAGCATCAATCACTTGTGTTGTCTAAAGCTTGTCTCCAGTGCAAAACAG 475
Qy 241 TATATTTGTGAGTGCATCTGAAAAGTGGAGAGATGTGCTCTCTGGAGTGTGATGTGCAGA 300
Db 476 TATATTTGTGAGTGCATCTGAAAAGTGGAGAGATGTGCTCTCTGGAGTGTGATGTGCAGA 535
Qy 301 TGTATTTGAATTTTACAAAATTTAGCTTGCACACATCTAGGCATACAGTTCTACCAAGTTCTCT 360
Db 536 TGTATTTGAATTTTACAAAATTTAGCTTGCACACATCTAGGCATACAGTTCTACCAAGTTCTCT 595
Qy 361 GTTGGGAATCAGCGAAGAAAGGCTTTTATGTCACAGGACATTAACCTGAAATCTCTGTT 420
Db 596 GTTGGGAATCAGCGAAGAAAGGCTTTTATGTCACAGGACATTAACCTGAAATCTCTGTT 655
Qy 421 GTGATGTCTACCAAGCTTGAAGTATTAATTAATCTCTTAGTATCAAGATATCACCAAGCTGG 480
Db 656 GTGATGTCTACCAAGCTTGAAGTATTAATTAATCTCTTAGTATCAAGATATCACCAAGCTGG 715
Qy 481 ATTAGCTCCATGAGTATTAATTCGATCCACCGAAACACAAGAGGACACAGTGGTAGCACTC 540
Db 716 ATTAGCTCCATGAGTATTAATTCGATCCACCGAAACACAAGAGGACACAGTGGTAGCACTC 775
Qy 541 TCGGTGACCTGCGATCCTGAAAGTCTGGATTGTTTACCTCGGAAATAGTCAATGCAATGCAAGAT 600
Db 776 TCGGTGACCTGCGATCCTGAAAGTCTGGATTGTTTACCTCGGAAATAGTCAATGCAATGCAAGAT 835
Qy 601 ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC 660
Db 836 ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC 895
Qy 661 TCTTTTTGTGCAATTTACACAAAGGTCACATTTTGTGTTGTTGTTCCAAAATATTGGAGGGTG 720
Db 896 TCTTTTTGTGCAATTTACACAAAGGTCACATTTTGTGTTGTTGTTCCAAAATATTGGAGGGTG 955
Qy 721 TTCGATGCGGAGACATTTCTTGTGTTGTTGAGGTCCTAGTGAAAATGACAGACATGG 780
Db 956 TTCGATGCGGAGACATTTCTTGTGTTGTTGTTGAGGTCCTAGTGAAAATGACAGACATGG 1015
Qy 781 ACCGGGGGGACCTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT 840
Db 1016 ACCGGGGGGGACCTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT 1075
Qy 841 TATATTTTACAAACTACCTCCAGTTGCTTCCAGCTAGTAGTATTCATTCGCGAGTGATGTG 900
Db 1076 TATATTTTACAAACTACCTCCAGTTGCTTCCAGCTAGTAGTATTCATTCGCGAGTGATGTG 1135
Qy 901 GGGAGGCGCAGTTGAAAATTTAAATTCCTCTGTGACAAATATCTCTCTGGATCGAAAAGAT 960
Db 1136 GGGAGGCGCAGTTGAAAATTTAAATTCCTCTGTGACAAATATCTCTCTGGATCGAAAAGAT 1195
Qy 961 AAAGAGTTGCTAAATTTGTCCTCTCTGTACTCTGGTTCTCTATGAGTGCAGAGATATTTTC 1020
Db 1196 AAAGAGTTGCTAAATTTGTCCTCTCTGTACTCTGGTTCTCTATGAGTGCAGAGATATTTTC 1254
Qy 1021 CATAAACTGTTAATTCAGGCTGATTTCTCTGGAAGGTGTAATTTTGGAAACATATCAGAC 1080
Db 1255 CATAAACTGTTAATTCAGGCTGATTTCTCTGGAAGGTGTAATTTTGGAAACATATCAGAC 1314
Qy 1081 ACAGCTGATAAAACAGGGAAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTGCAAA 1140
Db 1315 ACAGCTGATAAAACAGGGAAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTGCAAA 1374
Qy 1141 GAGGCAATTTGATTAACCTGAATCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1200
Db 1375 GAGGCAATTTGATTAACCTGAATCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1434
```


1201 CCATATAGTAACTCTTAAGTAACTGCAAGTGTGTACATACCAGCACATGAGCA 1260
1435 CCCAATAGTAACTCTTAAGTAACTGCAAGTGTGTACATACCAGCACATGAGCA 1494
1261 CTTGTTTGTGTCGTGAAGTGAAGCATAGTTATTTGCTGCTGCAACACAGAGCGCCATA 1320
1495 CTTGTTTGTGTCGTGAAGTGAAGCATAGTTATTTGCTGCTGCAACACAGAGCGCCATA 1554
1321 GTACAGCTGTTGCAAGGGGAACACATGCTCAGAAAGAGTTGGCCACCTCACAGAACACTC 1380
1555 GTACAGCTGTTGCAAGGGGAACACATGCTCAGAAAGAGTTGGCCACCTCACAGAACACTC 1614
1381 CGTGGTCATCGGAACCAAGTCAATGTTGCTATATCTCATCAGGTCCTCAGCTCGTAT 1440
1615 CGTGGTCATCGGAACCAAGTCAATGTTGCTATATCTCATCAGGTCCTCAGCTCG --- 1670
1441 GATCAAGATACCTGATATCTGGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1500
1671 ----- 1670
1501 TCTGGAGAAATGAACAATATCTTCTGTGTTTCATGGTGGTGAGATTACTCAACTTCTAGTT 1560
1671 ----- 1670
1561 CCACCTGAAACTGTAGTGAAGTACAGCATGCACTGCTCTGTAGCCAGTACCAC 1620
1671 -----GCAAGAGTACAGCATGCACTGCTCTGTAGCCAGTACCAC 1712
1621 TCAGTAGACTTCTTAAGTTTGCAGAGAAATAATGCAATATGTTGGCATCTCGTCACTT 1680
1713 TCAGTAGACTTCTTAAGTTTGCAGAGAAATAATGCAATATGTTGGCATCTCGTCACTT 1772
1681 TTTCCCTATTCAAGTAATCAAAATGGAGCCCTTCTGATGATTAACCTGTTGGTGGGTGTCA 1740
1773 TTTCCCTATTCAAGTAATCAAAATGGAGCCCTTCTGATGATTAACCTGTTGGTGGGTGTCA 1832
1741 GATGGTCTGTGTACGTCGGCAATGGATACCTGGTGCATTTGATGTTGTGTGATGGG 1800
1833 GATGGTCTGTGTACGTCGGCAATGGATACCTGGTGCATTTGATGTTGTGTGATGGG 1892
1801 ATAAACAGCAGTGTGATTTCTAAACGCTTGTGATGAAGCTGCTCTGCTGCTGTGATTCA 1860
1893 ATAAACAGCAGTGTGATTTCTAAACGCTTGTGATGAAGCTGCTCTGCTGCTGTGATTCA 1952
1861 CTTAGTCATCCAGCAGTCAACCTAAACAAAGCTATGACGAGAGTGTCTTGTGCTCTT 1920
1953 CTTAGTCATCCAGCAGTCAACCTAAACAAAGCTATGACGAGAGTGTCTTGTGCTCTT 2012
1921 AAAAATATGCCCCATCATAGCTACAAACCTTGGCAACTAACTCTTGGCTTCTGAGGCA 1980
2013 AAAAATATGCCCCATCATAGCTACAAACCTTGGCAACTAACTCTTGGCTTCTGAGGCA 2072
1981 TCTGACAAGGGAATTTTACCTAAATATCTCAAACTCCCTGATGTTCAAGCAATAAG 2040
2073 TCTGACAAGGGAATTTTACCTAAATATCTCAAACTCCCTGATGTTCAAGCAATAAG 2132
2041 ACAAACTTAACAGACCGGACATACATGTGCTATTTCTTGTGATGGAAGCGTTGATTAT 2100
2133 ACAAACTTAACAGACCGGACATACATGTGCTATTTCTTGTGATGGAAGCGTTGATTAT 2192
2101 CAATCTCTGATGAAGAGCTCTAGCCGAATACCTGCTCTTATTTTCCCGAGAGATTG 2160
2193 CAATCTCTGATGAAGAGCTCTAGCCGAATACCTGCTCTTATTTTCCCGAGAGATTG 2252
2161 CAATAAGCATCTGGCAGTTCAGACAAGGGGCTCTTTTAACTGGAACACGAGCA 2220
2253 CAATAAGCATCTGGCAGTTCAGACAAGGGGCTCTTTTAACTGGAACACGAGCA 2312
2221 GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCTTGTAT 2280
2313 GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCTTGTAT 2372
2281 GATGAAGAGAGGATGAGGAGATAATGAGGCGAGAGGGAAGAAAGTGTATCTGTAATAT 2340

RESULT 13

2373 GATGAAGAGAGGATGAGGAGATAATGAGCGAGAGGGAAGAAAGTATCTGTAATAT 2432
2341 CGGTCCAGCAAAATCAAAAGCCATTGACCCATTATAGAAATATAATTTAACTATGGACACTGCA 2400
2433 CGGTCCAGCAAAATCAAAAGCCATTGACCCATTATAGAAATATAATTTAACTATGGACACTGCA 2492
2401 AAGCTGTTATGCTCTGCTTCAAGCTGCGGGTTGAATGAAGTACTGGATGAAGTTG 2460
2493 AAGCTGTTATGCTCTGCTTCAAGCTGCGGGTTGAATGAAGTACTGGATGAAGTTG 2552
2461 CTGGATGCGCTTGGAAATGCTGAAACCCCACTGCAAGTATCGTTTGGCCCTCTGTGCAAGA 2520
2553 CTGGATGCGCTTGGAAATGCTGAAACCCCACTGCAAGTATCGTTTGGCCCTCTGTGCAAGA 2612
2521 GAGAGCCCATATGCTCACTGATGCTGCGGGTTAATCAGCGCTCTGTAAGTGTACAT 2580
2613 GAGAGCCCATATGCTCACTGATGCTGCGGGTTAATCAGCGCTCTGTAAGTGTACAT 2672
2581 GGGAAAACAGAACTAGAAAGGAGCTGCCAGCGTCTGAGGAGTAGGAAAGGGAACCTTAC 2640
2673 GGGAAAACAGAACTAGAAAGGAGCTGCCAGCGTCTGAGGAGTAGGAAAGGGAACCTTAC 2732
2641 GAGTGTCCGCTGCGCTCACCACAGCATCTCTCTCTATCATTTCTTTGGCAATACT 2700
2733 GAGTGTCCGCTGCGCTCACCACAGCATCTCTCTCTATCATTTCTTTGGCAATACT 2792
2701 TTAATCAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAGGCTCTTACCAGG 2760
2793 TTAATCAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAGGCTCTTACCAGG 2852
2761 CCACCTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTCCCTCCCACTTCCAGT 2820
2853 CCACCTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTCCCTCCCACTTCCAGT 2912
2821 AATATTTGTCNAGGACAGATTAACAAGTTGCTGACCTGTGCTTTCCGCTCGTCTGAT 2880
2913 AATATTTGTCNAGGACAGATTAACA --- 2937
2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTCTGCTTTTACATACCTGTTCTTATGTA 2940
2938 ----- 2937
2941 AATGAAGTTGGAGTCAGTTAGTGTCTATGCACTGTGTTATGCTGCCAGACCTTACTGGGA 3000
2938 ---AAGTTGGAGTCAGTTAGTGTCTATGCACTGTGTTATGCTGCCAGACCTTACTGGGA 2993
3001 TTGGATAAATTTAGGCTCTCCCTTCTGGAGATGCTGCGCCGGAAGATGGCAAGATCGATGC 3060
2994 TTGGATAAATTTAGGCTCTCCCTTCTGGAGATGCTGCGCCGGAAGATGGCAAGATCGATGC 3053
3061 TTGGAGGTGAGAGAGCCGACAGGCTCTGCTGCGGAACTGAGAGAAATTTGAGCAG 3120
3054 TTGGAGGTGAGAGAGCCGACAGGCTCTGCTGCGGAACTGAGAGAAATTTGAGCAG 3113
3121 GCAGGCAAGAGAGCCCAATGATGCTTGGCTCTTACTTACTCTCAGTACATAGACCCAC 3180
3114 GCAGGCAAGAGAGCCCAATGATGCTTGGCTCTTACTTACTCTCAGTACATAGACCCAC 3173
3181 GTCATATCACTGGAGTCAATCAGAGCCGCGCAGACTATCACCAGGCTCTGATGCC 3240
3174 GTCATATCACTGGAGTCAATCAGAGCCGCGCAGACTATCACCAGGCTCTGATGCC 3233
3241 TCAGGSCCTGAAGCAAAAGTCCAGGAGGAGAGATGACCTGTTTCCAGCATGATCACC 3300
3234 TCAGGSCCTGAAGCAAAAGTCCAGGAGGAGAGATGACCTGTTTCCAGCATGATCACC 3293
3301 ACTGTTGCTTATCAAGTGTCCC 3323
3294 ACTGTTGCTTATCAAGTGTCCC 3316

AK057141	3316 bp	mRNA	linear	PRI 30-JAN-2004
LOCUS				
DEFINITION	Homo sapiens cDNA FLJ32579 fis, clone SPLEN2000267, highly similar to Rattus norvegicus TGF-beta resistance-associated protein (TRAG) mRNA.			
ACCESSION	AK057141			
VERSION	AK057141.1	GI:16552730		
WORDS	oligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakematsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oabayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirao, K., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Konai, P., Hara, R., Hara, A., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, K., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs			
JOURNAL	Nat. Genet.	36	(1), 40-45	(2004)
PUBLISHED	14702039			
REFERENCE	2			
AUTHORS	Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Iehii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakami, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.			
TITLE	NEDO human cDNA sequencing project			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 3316)			
AUTHORS	Isogai, T., Otsuki, T. and Sugiyama, T.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.			
FEATURES	Location/Qualifiers			
source	1. .3316 /organism="Homo sapiens"			

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SPLEN2000267"
/tissue_type="spleen"
/clone_lib="SPLEN2"
/note="cloning vector: pME18SFL3"

ORIGIN

Query Match	62.8%	Score	2807.4;	DB 9;	Length	3316;
Best Local Similarity	92.7%	Pred. No.	0;			
Matches	3080;	Conservative	0;	Mismatches	1;	Indels 242; Gaps 3;
QY	1	ATGGCAGGAAACAGCGCTTGTCTACCCCAATTTGTTCTTTGGGGTCGAAAAGCGCCACACAT	60			
Db	236	ATGGCAGGAAACAGCGCTTGTCTACCCCAATTTGTTCTTTGGGGTCGAAAAGCGCCACACAT	295			
QY	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGACGATCGTAAACAGGATGTCCAGAC	120			
Db	296	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGACGATCGTAAACAGGATGTCCAGAC	355			
QY	121	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAAGCTGCAAAATTAATCTCGAGCATGTTG	180			
Db	356	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAAGCTGCAAAATTAATCTCGAGCATGTTG	415			
QY	181	TTTGGTTCATACAGCATCAATCACTTGTGTCTAAAGCTTGTGCTTCCAGTGCACAAACAG	240			
Db	416	TTTGGTTCATACAGCATCAATCACTTGTGTCTAAAGCTTGTGCTTCCAGTGCACAAACAG	475			
QY	241	TATATTGTGATGTCATCTGAAAGTGGAGAGATGTGCCTCTGGATGTGATGATGGCAGA	300			
Db	476	TATATTGTGATGTCATCTGAAAGTGGAGAGATGTGCCTCTGGATGTGATGATGGCAGA	535			
QY	301	TGTATTGAATTTACAAATTTAGTTCGCACACATCTGGCATACAGTTCTTACCAGTTCTCT	360			
Db	536	TGTATTGAATTTACAAATTTAGTTCGCACACATCTGGCATACAGTTCTTACCAGTTCTCT	595			
QY	361	GTTCGGAAATCAGCAGAGAGAGGCTTTTATGCCACGACATATACCTCGAAATTCCTTGT	420			
Db	596	GTTCGGAAATCAGCAGAGAGAGGCTTTTATGCCACGACATATACCTCGAAATTCCTTGT	655			
QY	421	GTGGATGTCTACAGCGCTTGAAGTATATATCTCTTAGTATCAAGATATCAACAGCTGG	480			
Db	656	GTGGATGTCTACAGCGCTTGAAGTATATATCTCTTAGTATCAAGATATCAACAGCTGG	715			
QY	481	ATTAGTCCATCAGTATTTATTCGATCCACCGAACACAGAGAGACACAGTGTGACATC	540			
Db	716	ATTAGTCCATCAGTATTTATTCGATCCACCGAACACAGAGAGACACAGTGTGACATC	775			
QY	541	TCGGTGACTGGCATCTCTGAAGGCTTGGATTGTGTACCTCGAAATTAAGTGACATGCAAGAT	600			
Db	776	TCGGTGACTGGCATCTCTGAAGGCTTGGATTGTGTACCTCGAAATTAAGTGACATGCAAGAT	835			
QY	601	ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTTGTGATGATGTCAGAAATTC	660			
Db	836	ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTTGTGATGATGTCAGAAATTC	895			
QY	661	TCCTTTTGTGCAATTTACACAAAGGTCACCTTTTGTGTGTGTTCCAAATATTGGAGGGTG	720			
Db	896	TCCTTTTGTGCAATTTACACAAAGGTCACCTTTTGTGTGTGTTCCAAATATTGGAGGGTG	955			
QY	721	TTTCGATGCGGAGACATTTCTTGTGTCTTCAGGTCCTAGTGAAATTCGACAGACATGG	780			
Db	956	TTTCGATGCGGAGACATTTCTTGTGTCTTCAGGTCCTAGTGAAATTCGACAGACATGG	1015			
QY	781	ACCGGGGGGAGCTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT	840			
Db	1016	ACCGGGGGGAGCTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT	1075			
QY	841	TATATTACAAACTACCTGCCAGTTGCCCTTCCAGCTAGTGATTCATTCCGAGTGTATGG	900			
Db	1076	TATATTACAAACTACCTGCCAGTTGCCCTTCCAGCTAGTGATTCATTCCGAGTGTATGG	1135			
QY	901	GGGAAGGCAGTTTGAAAAATTTAAATTCCTCCTGTACACATATCTCTTGGATCGAAAAAGAT	960			

1136 GGGNAGGCGAGTTGAAATTTTAAATTCCTCTGTACACATATCTCTTTGGATCGAAAGAT 1195
QY
961 AAGAGGTTGCTAAATTTGCTCTCTCTGTTACTCGGTTCTTCTATGATGACGAGAAATTTTC 1020
Db
1196 AAGAGGTTGCTAAATTTGCTCTCTCTGTTACTCGG-TCTTCTATGATGACGAGAAATTTTC 1254
QY
1021 CATAACTGTTAAATTCAGGCTGATCTCTCTGGAAGGTTGAATATTGGAAATATCAGAC 1080
Db
1255 CATAACTGTTAAATTCAGGCTGATCTCTCTGGAAGGTTGAATATTGGAAATATCAGAC 1314
QY
1081 ACAGCTGATAAACAGGAGTGAAGAGGCTGCAATGACAACTTCTATTAGTTTGCAG 1140
Db
1315 ACAGCTGATAAACAGGAGTGAAGAGGCTGCAATGACAACTTCTATTAGTTTGCAG 1374
QY
1141 GAGGCAATTCATAACTGAAATCTTGTCTCTGGAATTTAGATCATCAGCTGAGTGTGAT 1200
Db
1375 GAGGCAATTCATAACTGAAATCTTGTCTCTGGAATTTAGATCATCAGCTGAGTGTGAT 1434
QY
1201 CCCAATAGTAAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGACATGAGCGA 1260
Db
1435 CCCAATAGTAAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGACATGAGCGA 1494
QY
1261 CTGTTGTTGCTGCTGAAGTGAAGCATAGTTATTGTACCTGCCACACAGACGGCCATA 1320
Db
1495 CTGTTGTTGCTGCTGAAGTGAAGCATAGTTATTGTACCTGCCACACAGACGGCCATA 1554
QY
1321 GTACAGCTGTTGCAAGGGGAACCATGCTCAGAAAGGTTGGCCACCTCACAAGAACATC 1380
Db
1555 GTACAGCTGTTGCAAGGGGAACCATGCTCAGAAAGGTTGGCCACCTCACAAGAACATC 1614
QY
1381 CGTGGTCACTCGGAACAAAGTCAATGTTGCTATATCTCATCAGTCTCAGCTCGGTAT 1440
Db
1615 CGTGGTCACTCGGAACAAAGTCAATGTTGCTATATCTCATCAGTCTCAGCTCG- 1670
QY
1441 GATCAAAGATACCTGATATCTGAGGTGTGGAATTTTTCAGTCATATTTGGGACATATT 1500
Db
1671 ----- 1670
QY
1501 TCTGGAGAAATGAAACATATCTTCTGTGTTCTAGTGTGAGATTACTCAACTTCTAGTT 1560
Db
1671 ----- 1670
QY
1561 CCACCTGAAACCTGTAGTGAAGAGTACAGCATGCACTGCTGTAGCCAGTGAACAC 1620
Db
1671 -----GCAAGATACAGCATGCACTGCTGTAGCCAGTGAACAC 1712
QY
1621 TCAGTAGGACTTCAAGTTTGCAGAGAAATGATATGTTGGCATCTGCTCACTT 1680
Db
1713 TCAGTAGGACTTCAAGTTTGCAGAGAAATGATATGTTGGCATCTGCTCACTT 1772
QY
1681 TTTCTTATTCAAGTAAATCAAAATGGAGGCTTCTGATGATTAACCTGGTGGTGGGTTCTCA 1740
Db
1773 TTTCTTATTCAAGTAAATCAAAATGGAGGCTTCTGATGATTAACCTGGTGGTGGGTTCTCA 1832
QY
1741 GATGGTCTGTGACGCTGCGGAAATGGATPACTGGTGCAATGGATCGTTGTGTGATGGGG 1800
Db
1833 GATGGTCTGTGACGCTGCGGAAATGGATPACTGGTGCAATGGATCGTTGTGTGATGGGG 1892
QY
1801 ATAAACAGCTGTGAGATTCTTAAAGCTTGTGATGAAGCTGTTCTGCTGCTGTGATTCA 1860
Db
1893 ATAAACAGCTGTGAGATTCTTAAAGCTTGTGATGAAGCTGTTCTGCTGCTGTGATTCA 1952
QY
1861 CTTAGTCATCCAGCAGTCAACTTAAACAGCTTATGACGAGAGCTAGTCTTGTGCTCTT 1920
Db
1953 CTTAGTCATCCAGCAGTCAACTTAAACAGCTTATGACGAGAGCTAGTCTTGTGCTCTT 2012
QY
1921 AAAAATATGCCCATCATAGCTACAAACCTTTGCAACTTAACTCTTGGCTTCTGAGGCA 1980
Db
2013 AAAAATATGCCCATCATAGCTACAAACCTTTGCAACTTAACTCTTGGCTTCTGAGGCA 2072
QY
1981 TCTGACAGGGAAATTTACCTAAATTTCTCATATCTCCCTGATGTTGTTCAAGCAATAAAG 2040

2073 TCTGACAGGGAAATTTTACCTAAATATTCTCATAACTCCCTGATGTTCAAGCAATAAAG 2132
QY
2041 ACAAACCTTAACAGACCCGACATACATGTGCTATTCTTTGATGTGAAGCGTTGATTAAT 2100
Db
2133 ACAAACCTTAACAGACCCGACATACATGTGCTATTCTTTGATGTGAAGCGTTGATTAAT 2192
QY
2101 CAACTCTGACTGAAGAAAGCCTCTAGGCCGAATACCTCTTATTTCCTCCAGAGAAATTTG 2160
Db
2193 CAACTCTGACTGAAGAAAGCCTCTAGGCCGAATACCTCTTATTTCCTCCAGAGAAATTTG 2252
QY
2161 CAAAAGAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTAACTCGGAAACCGAGCAGCA 2220
Db
2253 CAAAAGAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTAACTCGGAAACCGAGCAGCA 2312
QY
2221 GTTCTCTTCCAAAGAGTGAAGAAACGATCAAGAGAAACATCAAGAGAACCTCTCTTGAT 2280
Db
2313 GTTCTCTTCCAAAGAGTGAAGAAACGATCAAGAGAAACATCAAGAGAACCTCTCTTGAT 2372
QY
2281 GATGAAGAGAGGATGAGGAGATTAATGAGCAGAGAGGAGAGAAAGTGTATCTGATAT 2340
Db
2373 GATGAAGAGAGGATGAGGAGATTAATGAGCAGAGAGGAGAGAAAGTGTATCTGATAT 2432
QY
2341 CGGTCCAGCAAAATCAAGAGCATTTGACCTATTAGATAATAATTTTAACTATGAGACATGCA 2400
Db
2433 CGGTCCAGCAAAATCAAGAGCATTTGACCTATTAGATAATAATTTTAACTATGAGACATGCA 2492
QY
2401 AAGCTGTTTATGCTCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTGAGTGAAGTTTC 2460
Db
2493 AAGCTGTTTATGCTCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTGAGTGAAGTTTC 2552
QY
2461 CTGGATCGCTTGGAAATGCTGAAACCCCATGCAACCGTATCGTTTGGCTCTTGTGCAAGA 2520
Db
2553 CTGGATCGCTTGGAAATGCTGAAACCCCATGCAACCGTATCGTTTGGCTCTTGTGCAAGA 2612
QY
2521 GAGGGCCATATGTCAGTCTGATGCTGCGGGTTAATACAGCTGCTTGTAACTGTGCACAT 2580
Db
2613 GAGGGCCATATGTCAGTCTGATGCTGCGGGTTAATACAGCTGCTTGTAACTGTGCACAT 2672
QY
2581 GGGAAACAGAAAGTAGAAAGAGCTGCCAGCGTCTGAGGGAGTAGGAAAGGAACTTAC 2640
Db
2673 GGGAAACAGAAAGTAGAAAGAGCTGCCAGCGTCTGAGGGAGTAGGAAAGGAACTTAC 2732
QY
2641 GAGATGTCCTGCTGCTGCTACCAACAGATCTCTGCTATCATTTCTTTGGCAAAATACT 2700
Db
2733 GAGATGTCCTGCTGCTGCTACCAACAGATCTCTGCTATCATTTCTTTGGCAAAATACT 2792
QY
2701 TTAATGAGTATGACCAATGCAACTTTTATTGGTGATCATATGAAGAGGCTCTACGAG 2760
Db
2793 TTAATGAGTATGACCAATGCAACTTTTATTGGTGATCATATGAAGAGGCTCTACGAG 2852
QY
2761 CCACCTTAGACCAAGCACCCAGACCTTTCTAAGGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
Db
2853 CCACCTTAGACCAAGCACCCAGACCTTTCTAAGGGCAAGGGGTTCCCTCCAACTTCCAGT 2912
QY
2821 AATATTGTGCAAGGACAGATTAAACAAAGTTGTGCAACCTGTGCTGTTTCCGCTCGGTCTGAT 2880
Db
2913 AATATTGTGCAAGGACAGATTAAAC----- 2937
QY
2881 GCTGATCACTCTGCGCTCTGACCCCTCTCTGCTCTGCTTTTACATACCTGTTTCTTAGTA 2940
Db
2938 ----- 2937
QY
2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGTGTCGACAGACCTTACTGGGA 3000
Db
2938 -----AAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGTGTCGACAGACCTTACTGGGA 2993
QY
3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCCGGAGAGTGGCAAGATCGATGC 3060
Db
2994 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCCGGAGAGTGGCAAGATCGATGC 3053
QY
3061 TTGGAGGTGAGAGAGCCGACAGCCCTGCTCTGCGGGAATCTGAGAGAAATTTGAGCAG 3120
Db
3054 TTGGAGGTGAGAGAGCCGACAGCCCTGCTCTGCGGGAATCTGAGAGAAATTTGAGCAG 3113

||||| 2842 ACCACATAGTAAAGAGTCCACACACACACGCGCTTCGACGACACCCAGTCCAC 2901
||||| 3874 CAGAAATATGACACAAACATCTTTGACGAGCTAAAGGGGAAATTTTGAGAGTCATTGAA 3933
||||| 2902 CAGAGTATACACACACCAACCTTGGCGAGGGCTAAAGGGGAAATCTTGAGAGTCATTGAA 2961
||||| 3934 ATTCTTATTGAAAGATGCCACAGATGTTGGATCTTCTCGTGAGGTTATGGACATC 3993
||||| 2962 ATTCTCATTTGAAAGATGCCACGAGTGGTGGATCTTCTTGAGAGTGGATGATC 3021
||||| 3994 ATTATGCTACTGCTTGAAGATCTTTAGTTTAAAGAAAGAGGCTTTCAGAAATGTTTCCCA 4053
||||| 3022 ATCATGCTACTGCTGGAAGATCTTTAGTGAAGAAAGAGGCTTTCAGGAGTGTTCCTCA 3081
||||| 4054 GCATCTGACAGGTTCTACATGTCAGTATTTATGACGCGGAATCACAGAATAGCAGTTGGA 4113
||||| 3082 GCTATCTGACAGGTTCTACATGTCAGTATTTATGACGCGGAGTTCACAGAATCGCAGTTGGA 3141
||||| 4114 GCTCGCATGCTTCACTGCGGCTGTACGATCCGACCTGGAATGCGAAATGTCAGACATCCAT 4173
||||| 3142 GCACGCGCATGCTTCACTGCGGCTGTATGACATCCGACCTGGAATGTCAGACATCCAT 3201
||||| 4174 GGACACAGGAGCAATCACTGCGGCTGTATGCTTCTGATGGAAGATATCTTGCCACC 4233
||||| 3202 GGGACACAGGAGCAATCACTGCGGCTGTATGCTTCTGATGCGGCTTACCTTGCCACC 3261
||||| 4234 TACTCAAACTGACAGCCCAATTTCTTTTGGCAGATGAACACGTCATCTGCGACT 4293
||||| 3262 TACTCGAACCACGACGACCAATTTCTTCTGCGCAGATGAACACCTCACTCTCTGGGAGC 3321
||||| 4294 ATCGCATGCTGAATCGGACCTTCACTGCGGCTGTATGCTTCTGATGCGGCTTACCTTGCCACC 4353
||||| 3322 ATTTGGAATGCTGAATCACTGCGGCTGTATGCTTCTGATGCGGCTTACCTTGCCACC 3381
||||| 4354 GTGACGCGGCTGCGGCTGCCCAATGCGGCTTCACTGCGGCTTCACTGCGACT 4413
||||| 3382 GTGACGCGGCTGCGGCTGCCCAATGCGGCTTCACTGCGGCTTCACTGCGACT 3441
||||| 4414 TCCAAACGCAATCTCACTGCGGCTTCACTGCGGCTTCACTGCGGCTTCACTGCGACT 4473
||||| 3442 TCCAAACGCAATCTCACTGCGGCTTCACTGCGGCTTCACTGCGGCTTCACTGCGACT 3501

RESULT 15
BC050352
LOCUS
DEFINITION Homo sapiens WD repeat domain 7, mRNA (cdna clone IMAGE:5265024), partial cds.
ACCESSION BC050352
VERSION BC050352.1 GI:30048099
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 4460)
Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.L., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Tashiro, H., Tashiro, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analyses of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 4460)
Strausberg, R.
Direct Submission
Submitted (08-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAC Plate: 90 Row: p Column: 18.
Location/Qualifiers
1..4460
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5265024"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
<1..4460
/gene="WDR7"
/note="synonyms: KIAA0541, TRAG"
/db_xref="LocusID:23335"
<1..1845
/gene="WDR7"
/codon_start=1
/product="WDR7 protein"
/protein_id="AAH50352.1"
/db_xref="GI:30048100"
/db_xref="LocusID:23335"
/translation="KGTGVSRVAVTQHLSTIIISLNTLMSTNATFTGDHMKKGFT
PERSTPLSKARGSPFTSSNIVQVQIKQVAVSVASDADSDSDPSAPALHTCF
LVNCGSOLAAMHCVMLDGLDFRPFLLEMLARWRQDCLREAAQALLAEHL
RTEQGRKEDADAWAPYLPOYIDHIVISQVTSSEAAQTITPADSGPAKQVEEHL
VDDDTTCGLSSVPOMKISTSYERRKQATVILVIGVAGFAGABIEPKLLTRPS
SQIPEGSLVSGSNYSARHTKALTFLLOPPSPKLPHPSTIRRTALDILGRFT
VNEPVGSLVMSGLLELACADKQLANTMGPLSPADRSARSHALSILATARP
AFITTIKEVHRHTALANTQQNNMTTTLARAKEILRVEILIERKMTDVLV
EPTMDIIMYCLEGLVKKGLQCFPAICFTFMVSYERNHRIAVGARHGSVALYDIT
GKCTIIGHKGPITAVAFAPDGRYLATYSNTDSHISFWOMNTSLGSLGMLNSAPQLR
CKITQVQPPQSPSPSHNALKLARLIWTSNRNVILMAHDGKEHFMV"
1441..1650
/gene="WDR7"
/note="WD40; Region: WD40 domain, found in a number of eukaryotic proteins that cover a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly"
/db_xref="CDD:cd00200"

ORIGIN

Query Match	41.2%	Score	1841.8	DB	9	Length	4460
Best Local Similarity	99.9%	Prod. No.	0				
Matches	1843	Conservative	0	Mismatches	2	Indels	0
Gaps	0						
Qy	2629	AAGGGA	ACTTACGAGAGTGC	CCGTCGCTACACACAGCATCTCC	TGCTATCATTTCT	2688	
Db	1	AAGGGA	ACTTACGAGAGTGC	CCGTCGCTACACACAGCATCTCC	TGCTATCATTTCT	60	
Qy	2689	TTGCAAA	TACTTTAATGAGTATGACCAATGCAACTTTTAT	TGCTGATCATATGAAGAAG	2748		
Db	61	TTGCAAA	TACTTTAATGAGTATGACCAATGCAACTTTTAT	TGCTGATCATATGAAGAAG	120		
Qy	2749	GGTCT	TACGAGCCACCTAGACCAAGCACCCAGACCTTTCTA	AGGCAAGGGTTC	2808		
Db	121	GGTCT	TACGAGCCACCTAGACCAAGCACCCAGACCTTTCTA	AGGCAAGGGTTC	180		
Qy	2809	CCAACTT	CCAGTAATTTGTGCAAGGACAGATTAAACAAGTTGCTG	CACCTGTGCTTCC	2868		
Db	181	CCAACTT	CCAGTAATTTGTGCAAGGACAGATTAAACAAGTTGCTG	CACCTGTGCTTCC	240		
Qy	2869	GCTCG	GTCTGATGCTCACTCTGGCTCTGACCTCTCTCTGCTCTGCT	TCTTACATACC	2928		
Db	241	GCTCG	GTCTGATGCTCACTCTGGCTCTGACCTCTCTCTGCTCTGCT	TCTTACATACC	300		
Qy	2929	TGTTT	CTTAGTAATGAAGTTGGAGTCAGTTAGCTGTATGCACTGTGTTAT	TGCTGCCA	2988		
Db	301	TGTTT	CTTAGTAATGAAGTTGGAGTCAGTTAGCTGTATGCACTGTGTTAT	TGCTGCCA	360		
Qy	2989	GACCT	ACTGGGATGGATTAATTTAGGCTCCCTCTCGGAGATGCTGGCCCGAAGATG	3048			
Db	361	GACCT	ACTGGGATGGATTAATTTAGGCTCCCTCTCGGAGATGCTGGCCCGAAGATG	420			
Qy	3049	CAAGAT	CCATGCTTGGAGGTGAGAGAGCCGACAGGCCCTCTCTCGGCGGAAC	TGAGA	3108		
Db	421	CAAGAT	CCATGCTTGGAGGTGAGAGAGCCGACAGGCCCTCTCTCGGCGGAAC	TGAGA	480		
Qy	3109	AGAAT	TGAGCAGGACGAGAGAAAGCAATGATGCTGGGCTCTTTACTTAC	CTCAG	3168		
Db	481	AGAAT	TGAGCAGGACGAGAGAAAGCAATGATGCTGGGCTCTTTACTTAC	CTCAG	540		
Qy	3169	TACAT	AGACACGTCATATCACCTGGAGTCACATCAGAAGCCGCGAGATATCACCAG	3228			
Db	541	TACAT	AGACACGTCATATCACCTGGAGTCACATCAGAAGCCGCGAGATATCACCAG	600			
Qy	3229	GCTCT	GATGCTCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGAC	3288			
Db	601	GCTCT	GATGCTCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGAC	660			
Qy	3289	GATGA	CATACCACTGCTTATCAAGTGTGCCACAAATGAAAAAATTTCTACATCT	3348			
Db	661	GATGA	CATACCACTGCTTATCAAGTGTGCCACAAATGAAAAAATTTCTACATCT	720			
Qy	3349	TACGAG	AAAGACGGAAGCAAGTACCGCTATTTGTTTACTTGGAGTAAATAGGAGCTGAA	3408			
Db	721	TACGAG	AAAGACGGAAGCAAGTACCGCTATTTGTTTACTTGGAGTAAATAGGAGCTGAA	780			
Qy	3409	TTTGG	TCTGAAATTTGAACCTCTTAAACTATTTGACAGACCTGAGAGCTCTAGCAAAAT	3468			
Db	781	TTTGG	TCTGAAATTTGAACCTCTTAAACTATTTGACAGACCTGAGAGCTCTAGCAAAAT	840			
Qy	3469	CTTGAG	GGATTCGGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATCTTGC	3528			
Db	841	CTTGAG	GGATTCGGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATCTTGC	900			
Qy	3529	AAGCA	CTGAGTTTCTTGCTACAGCTCCAAAGCCCAAACTTCTCCACACAGCACT	3588			
Db	901	AAGCA	CTGAGTTTCTTGCTACAGCTCCAAAGCCCAAACTTCTCCACACAGCACT	960			
Qy	3589	ATCCG	AGAACAGCCATTTGATTTGAGAGTGGGTTCACTGTTGGGAGCCTTATCATG	3648			
Db	961	ATCCG	AGAACAGCCATTTGATTTGAGAGTGGGTTCACTGTTGGGAGCCTTATCATG	1020			
Qy	3649	GATGT	GTCGCTGTTCTGATGGGGCTTCTCGAACTTTGTGCGGATGCCGAGAAACAAC	3708			

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 03:39:26 ; Search time 2385 seconds
(without alignments)
11763.304 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcaggaacagccttgt.....agcaccgtctatggtctaa 4473

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4473	100.0	4473	21	US-10-645-335-1
2	4460.2	99.7	4517	17	US-10-093-463-153
3	3497	78.2	6224	17	US-10-120-988-91
4	2807.4	62.8	3316	17	US-10-094-749-1317
5	474.6	10.6	2578	17	US-10-108-260A-1051
6	150.2	3.4	585	16	US-10-029-386-7862
7	147.2	3.3	227	16	US-10-029-386-21562

8	125.4	2.8	585	13	US-10-027-632-96119	Sequence 96119, A
9	125.4	2.8	585	17	US-10-027-632-96119	Sequence 96119, A
10	125.4	2.8	590	13	US-10-027-632-320926	Sequence 320926, A
11	125.4	2.8	590	17	US-10-027-632-320926	Sequence 320926, A
12	125.4	2.8	598	13	US-10-027-632-237407	Sequence 237407, A
13	125.4	2.8	598	17	US-10-027-632-237407	Sequence 237407, A
14	125.4	2.8	598	17	US-10-027-632-237407	Sequence 237407, A
15	125.4	2.8	598	17	US-10-027-632-237407	Sequence 237407, A
16	103.8	2.3	501	16	US-10-029-386-8422	Sequence 8422, A
17	99	2.2	99	16	US-10-029-386-8422	Sequence 8422, A
18	88.4	2.0	555	16	US-10-029-386-8040	Sequence 8040, A
19	84	1.9	84	16	US-10-029-386-21740	Sequence 21740, A
20	57	1.3	65	10	US-09-908-975-3606	Sequence 3606, A
21	57	1.3	572	13	US-10-027-632-192215	Sequence 192215, A
22	57	1.3	572	17	US-10-027-632-192215	Sequence 192215, A
23	41.8	0.9	852	17	US-10-369-493-42703	Sequence 42703, A
24	41.4	0.9	786	19	US-10-437-963-19710	Sequence 19710, A
25	41.2	0.9	1007	20	US-10-425-115-1650	Sequence 1650, A
26	40.8	0.9	1159	17	US-10-398-221-1331	Sequence 1331, A
27	40.8	0.9	1995	17	US-10-398-221-3359	Sequence 3359, A
28	40.6	0.9	774	18	US-10-287-971-71	Sequence 71, Appl
29	39.6	0.9	514	16	US-10-029-386-4824	Sequence 4824, A
30	39.6	0.9	678	18	US-10-425-114-10913	Sequence 10913, A
31	39.6	0.9	701	18	US-10-424-599-89082	Sequence 89082, A
32	39.6	0.9	861	18	US-10-424-599-89104	Sequence 89104, A
33	39.6	0.9	1076	18	US-10-424-599-130835	Sequence 130835, A
34	39.6	0.9	312477	19	US-10-317-883A-12	Sequence 12, Appl
35	39	0.9	177556	10	US-09-952-213D-6	Sequence 6, Appl
36	38.8	0.9	1583	18	US-10-425-114-2214	Sequence 2214, A
37	38.8	0.9	2762	20	US-10-425-115-162201	Sequence 162201, A
38	38.8	0.9	67883	22	US-10-737-082-66	Sequence 66, Appl
39	38.8	0.9	67883	22	US-10-765-790-66	Sequence 66, Appl
40	38.6	0.9	230	9	US-09-867-701-9505	Sequence 9505, A
41	38.4	0.9	1694	20	US-10-425-115-38361	Sequence 38361, A
42	38.2	0.9	6482	8	US-08-781-986A-52	Sequence 52, Appl
43	38.2	0.9	6482	18	US-10-329-624-52	Sequence 52, Appl
44	38	0.8	902	18	US-10-424-599-59957	Sequence 59957, A
45	38	0.8	1170	17	US-10-282-122A-34230	Sequence 34230, A

ALIGNMENTS

RESULT 1
US-10-645-335-1
; Sequence 1, Application US/10645335
; Publication No. US2005006430A1
; GENERAL INFORMATION:
; APPLICANT: Masakazu Takeuchi
; APPLICANT: Yoshiaki Takai
; TITLE OF INVENTION: Rabconnectin-3-binding protein
; FILE REFERENCE: 2003946-0055
; CURRENT APPLICATION NUMBER: US/10/645,335
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: JP 2002-319521
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 1
; LENGTH: 4473
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4470)
US-10-645-335-1

Query Match 100.0%; Score 4473; DB 21; Length 4473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCAGGAACAGCCTTGTTCTTACCCATTGTTCTTTGGGTCGAAAGCGCCACAT 60
DB 1 ATGGCAGGAACAGCCTTGTTCTTACCCATTGTTCTTTGGGTCGAAAGCGCCACAT 60

QY	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAACAGGATGTCAGGAC	120	QY	1141	GAGGCATTTGATAAACTGAATCCTTGTCTCTGCTGGAAATTATAGATCAGCTGAGTGTGATT	1200
Db	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAACAGGATGTCAGGAC	120	Db	1141	GAGGCATTTGATAAACTGAATCCTTGTCTCTGCTGGAAATTATAGATCAGCTGAGTGTGATT	1200
QY	121	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTTG	180	QY	1201	CCCAATAGTAAATGAACTCTTTAAAGTAACTGCAAGTGTGTACATACACAGCAATGAGCGA	1260
Db	121	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTTG	180	Db	1201	CCCAATAGTAAATGAACTCTTTAAAGTAACTGCAAGTGTGTACATACACAGCAATGAGCGA	1260
QY	181	TTTGGTCAATACAGCATCAATCACTTGTGTTGTCTTAAAGTGTGCTTCCAGTGACAAACAG	240	QY	1261	CTTGTGTTGGTTCGTGAGATGGAAGCAATGTTTATGTTACCTGCCACACAGACGGCCATA	1320
Db	181	TTTGGTCAATACAGCATCAATCACTTGTGTTGTCTTAAAGTGTGCTTCCAGTGACAAACAG	240	Db	1261	CTTGTGTTGGTTCGTGAGATGGAAGCAATGTTTATGTTACCTGCCACACAGACGGCCATA	1320
QY	241	TATATTGTGATGCTCTGAAAGTGGAGAGATGCTCTGGGATGTGAGTGATGCGAGA	300	QY	1321	GTACAGCTCTGTCGAAGGGGAAACACATGCTCAGAAGGTTGGCCACCTCACAGAAACATCTC	1380
Db	241	TATATTGTGATGCTCTGAAAGTGGAGAGATGCTCTGGGATGTGAGTGATGCGAGA	300	Db	1321	GTACAGCTCTGTCGAAGGGGAAACACATGCTCAGAAGGTTGGCCACCTCACAGAAACATCTC	1380
QY	301	TGTATTGAAATTTACAAATTTAGCTTGCACATACCTAGGCAATACAGTTCTACCAAGTTCTCT	360	QY	1381	CGTGGTCATCCGAAACAAAGTCAATGTTTGTCTATATCTCTCATCAGGCTCTCAGCTCGGTAT	1440
Db	301	TGTATTGAAATTTACAAATTTAGCTTGCACATACCTAGGCAATACAGTTCTACCAAGTTCTCT	360	Db	1381	CGTGGTCATCCGAAACAAAGTCAATGTTTGTCTATATCTCTCATCAGGCTCTCAGCTCGGTAT	1440
QY	361	GTTGGGAATCAGCGAGAGAAAGCTTTTATGCGACGGAATTAACCTCGAAATCCTTGT	420	QY	1441	GATCAAGATACCTGATATCTCGAGGTGGGATTTTTCAGTCAATATTTTGGGACATATTT	1500
Db	361	GTTGGGAATCAGCGAGAGAAAGCTTTTATGCGACGGAATTAACCTCGAAATCCTTGT	420	Db	1441	GATCAAGATACCTGATATCTCGAGGTGGGATTTTTCAGTCAATATTTTGGGACATATTT	1500
QY	421	GTGGATGCTACCAAGCTTGAAGTATTAATCTCTTAGTATCAAGATATCACCAGACTGG	480	QY	1501	TCTGGAGAAATGAAACATATCTTCTGTGTTTCATGTTGGTGGTGAAGTTACTCAACTTCTAGTT	1560
Db	421	GTGGATGCTACCAAGCTTGAAGTATTAATCTCTTAGTATCAAGATATCACCAGACTGG	480	Db	1501	TCTGGAGAAATGAAACATATCTTCTGTGTTTCATGTTGGTGGTGAAGTTACTCAACTTCTAGTT	1560
QY	481	ATTAGTCTCCATGATGATTTATTCGATCCCAACCGAACACAGAGGACACAGTGGTAGCACTC	540	QY	1561	CCACCTGAAATCTGTAGTGCAGAGTACAGCATGTCATCTGCTGTAGCCAGTGAACAC	1620
Db	481	ATTAGTCTCCATGATGATTTATTCGATCCCAACCGAACACAGAGGACACAGTGGTAGCACTC	540	Db	1561	CCACCTGAAATCTGTAGTGCAGAGTACAGCATGTCATCTGCTGTAGCCAGTGAACAC	1620
QY	541	TCGGTGTAGTGGCATTCCTGAAAGTCTGATTTGTACCTCGGAAATAGTGACATGCAAGAT	600	QY	1621	TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCAATATGTTGGCATCTCGTCACTTT	1680
Db	541	TCGGTGTAGTGGCATTCCTGAAAGTCTGATTTGTACCTCGGAAATAGTGACATGCAAGAT	600	Db	1621	TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCAATATGTTGGCATCTCGTCACTTT	1680
QY	601	ACTGAGCCAAATTTGAGGAGGATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC	660	QY	1681	TTTCTTATTTCAAGTAAATCAATGGAGCCCTTCTGATGATTACTGCTGGTGGGGTGTCA	1740
Db	601	ACTGAGCCAAATTTGAGGAGGATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC	660	Db	1681	TTTCTTATTTCAAGTAAATCAATGGAGCCCTTCTGATGATTACTGCTGGTGGGGTGTCA	1740
QY	661	TCCTTTTGTGATTTACAAAGTCTACTTTTGTGTTGTTGTTTCCAAATATTTGGAGGGTG	720	QY	1741	GATGTTTCTGTGTACGCTCTGGCAATGGAATATCTGGTGCATTTGGATCTGTTGTGTATGGG	1800
Db	661	TCCTTTTGTGATTTACAAAGTCTACTTTTGTGTTGTTGTTTCCAAATATTTGGAGGGTG	720	Db	1741	GATGTTTCTGTGTACGCTCTGGCAATGGAATATCTGGTGCATTTGGATCTGTTGTGTATGGG	1800
QY	721	TTGATGCCGAGACTATTCCTGTTGTGTTTCAGGTCTTAGTGAAATAGCAAGATG	780	QY	1801	ATAACAGAGTTGAGATTTCTAAGCTTGTGATGAGCTGTTCCCTGCTGCTGTGATTCA	1860
Db	721	TTGATGCCGAGACTATTCCTGTTGTGTTTCAGGTCTTAGTGAAATAGCAAGATG	780	Db	1801	ATAACAGAGTTGAGATTTCTAAGCTTGTGATGAGCTGTTCCCTGCTGCTGTGATTCA	1860
QY	781	ACGGGGGGGACTTTGTTCTCATCAGATAAGTCAATTTGGACAGAAATGGCCAAAGT	840	QY	1861	CTTAGTCTCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGTGCTCTT	1920
Db	781	ACGGGGGGGACTTTGTTCTCATCAGATAAGTCAATTTGGACAGAAATGGCCAAAGT	840	Db	1861	CTTAGTCTCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGTGCTCTT	1920
QY	841	TATATTACAACTACCTGCGAGTTGCTTCCAGCTAGTCAATTTCCCGAGTGTG	900	QY	1921	AAAAATATGGCCCATCATTAAGCTACAAACCTTTGCAACTAACTCTTGGCTTCTGAGGCA	1980
Db	841	TATATTACAACTACCTGCGAGTTGCTTCCAGCTAGTCAATTTCCCGAGTGTG	900	Db	1921	AAAAATATGGCCCATCATTAAGCTACAAACCTTTGCAACTAACTCTTGGCTTCTGAGGCA	1980
QY	901	GGGAAGCGAGTTGAAATTTAAATTCCTCTGTACAAATATTCCTTGGATCGAAAGAT	960	QY	1981	TCTGCAAGGGGAAATTTTACATAATTTCTCACTCCCTGATGTTTCAAGCAATAAAG	2040
Db	901	GGGAAGCGAGTTGAAATTTAAATTCCTCTGTACAAATATTCCTTGGATCGAAAGAT	960	Db	1981	TCTGCAAGGGGAAATTTTACATAATTTCTCACTCCCTGATGTTTCAAGCAATAAAG	2040
QY	961	AAAGAGTTGTAATTTGCTCTCTGTTACTCGGTTCTTCTATGGATCGAGAAATATTC	1020	QY	2041	ACAAACTAACAGACCCGGACATACATGTCATTCTTTGATGTGGAAGCTTGAATT	2100
Db	961	AAAGAGTTGTAATTTGCTCTCTGTTACTCGGTTCTTCTATGGATCGAGAAATATTC	1020	Db	2041	ACAAACTAACAGACCCGGACATACATGTCATTCTTTGATGTGGAAGCTTGAATT	2100
QY	1021	CATAAACTGTTAAATTCAGGGTGAATTTCTCTGGAGGTTGAATATTTTGGAAACATATCAGAC	1080	QY	2101	CAACTCCTGACTGAAGAGCCTCTAGCCGGAATACCTGCTCTTATTTTCCCGAGGAATTTG	2160
Db	1021	CATAAACTGTTAAATTCAGGGTGAATTTCTCTGGAGGTTGAATATTTTGGAAACATATCAGAC	1080	Db	2101	CAACTCCTGACTGAAGAGCCTCTAGCCGGAATACCTGCTCTTATTTTCCCGAGGAATTTG	2160
QY	1081	ACAGCTGATAAACAGGGAAGTGAAGAGGGCTGCAATGACAACTTCTATTTAGTTTGC	1140	QY	2161	CAAAAAGCATCTGGCAGTTTCAGAAAGGGGGCTCTTTTTTAACTCGAAAAACGAGCAGCA	2220
Db	1081	ACAGCTGATAAACAGGGAAGTGAAGAGGGCTGCGCAATGACAACTTCTATTTAGTTTGC	1140	Db	2161	CAAAAAGCATCTGGCAGTTTCAGAAAGGGGGCTCTTTTTTAACTCGAAAAACGAGCAGCA	2220
				QY	2221	GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAGACATCAAGGAAACACCTCTCTTGAT	2280

Db	2221	GTTCTCTTCCAAACAGTGGAAGAACAACATCAAGAGAACATCAGGAAACCTCCTTGTAT
Qy	2281	GATGAAGAGGAGTAGAGGAGATAAATGAGGCAGAGAGGAAGAAAGTAGTCCTGAATAT
Db	2281	GATGAAGAGGAGTAGAGGAGATAATGAGGCAGAGAGGAAGAAAGTAGTCCTGAATAT
Qy	2341	CGGTCCAGCAAATCAAGACCATTGACCCCTATTAGATATATTAATTAACTATGGACACTGCA
Db	2341	CGGTCCAGCAAATCAAGACCATTGACCCCTATTAGAAATATATTAATTAACTATGGACACTGCA
Qy	2401	AAGCTGTTTATGTCTGCCCTTCACGCCTGGGGTTTGAAATGAAGTACTTGATGAAGTTTCG
Db	2401	AAGCTGTTTATGTCTGCCCTTCACGCCTGGGGTTTGAATGAGTACTTGATGAAGTTTTCG
Qy	2461	CTGSAATCGCCTTGGAATPGCTGAAACCCCACTGCACCGGTATCGTTTGGCCCTCTTGTCAAGA
Db	2461	CTGSAATCGCCTTGGAATPGCTGAAACCCCACTGCACCGGTATCGTTTGGCCCTCTTGTCAAGA
Qy	2521	GGAGGCCATATGTCACTGATGCTGCCGGGTATTAATCAGGCTGCTGTAAACTGTGCACAT
Db	2521	GGAGGCCATATGTCACTGATGCTGCCGGGTATTAATCAGGCTGCTGTAAACTGTGCACAT
Qy	2581	GGGAAAAACAGATGAGGAAGAAAGCTGCCAGCGTCTGAGGAGTAGAGAAAGGAACTTAC
Db	2581	GGGAAAAACAGATGAGGAAGAAAGCTGCCAGCGTCTGAGGAGTAGAGAAAGGAACTTAC
Qy	2641	GGAGTGTCCGTGTCGCCGTACCACACAGCATCTCCTGTCTATCATTTCTTTGGCAATACT
Db	2641	GGAGTGTCCGTGTCGCCGTACCACACAGCATCTCCTGTCTATCATTTCTTTGGCAATACT
Qy	2701	TTAATGATGATGACCAATGCACTTTTATTTGGTGATCATATGAAGAAGGCTCTACGAG
Db	2701	TTAATGATGATGACCAATGCACTTTTATTTGGTGATCATATGAAGAAGGCTCTACGAG
Qy	2761	CCACCTAGACCAAGCACCCCAGACCTTTCTTAAAGCAAGGGTTCCCTCCCACTTCCAGT
Db	2761	CCACCTAGACCAAGCACCCCAGACCTTTCTTAAAGCAAGGGTTCCCTCCCACTTCCAGT
Qy	2821	AATATTGTGCAAGGACAGATTAACAAAGTTGTCTGCACCTGTGTGTTCGCTCGGTCTGAT
Db	2821	AATATTGTGCAAGGACAGATTAACAAAGTTGTCTGCACCTGTGTGTTCGCTCGGTCTGAT
Qy	2881	GCTGATCACTCTGGCTGTGACCCCTCCTTCTGTCTCTGCTTACATACCTGTTTCTTAGTA
Db	2881	GCTGATCACTCTGGCTGTGACCCCTCCTTCTGTCTCTGCTTACATACCTGTTTCTTAGTA
Qy	2941	AATGAAGTTTGGAGTCAGTTAGCTGTCTATGSCACTGTGTTATGCTGCGACACTTACGGGA
Db	2941	AATGAAGTTTGGAGTCAGTTAGCTGTCTATGSCACTGTGTTATGCTGCGACACTTACGGGA
Qy	3001	TTGGATAAATTTTAGGCTCCCTTCTGGAGATCTGGCCCGAAGATGGAAGATCGATGC
Db	3001	TTGGATAAATTTTAGGCTCCCTTCTGGAGATCTGGCCCGAAGATGGAAGATCGATGC
Qy	3061	TTGGAGGTGAGAGAGCCGACAGGCCCTGCTTCTGCGGAACTGAGAGAATTTGAGCAG
Db	3061	TTGGAGGTGAGAGAGCCGACAGGCCCTGCTTCTGCGGAACTGAGAGAATTTGAGCAG
Qy	3121	GCAGGAGGAAGGAAGCCATTGATGCTGGGCTCTTACTTACCTCAGTATCATAGACCAC
Db	3121	GCAGGAGGAAGGAAGCCATTGATGCTGGGCTCTTACTTACCTCAGTATCATAGACCAC
Qy	3181	GT CATATCACTTGAGTCA CAT CAGAAGCGCGCAGACTATCACACGGCTCCTGATGCC
Db	3181	GT CATATCACTTGAGTCA CAT CAGAAGCGCGCAGACTATCACACGGCTCCTGATGCC
Qy	3241	TCAGGGCTGAGCAAAAGTCTCAGGAGGAAGACATGACCTTGTGACGATGACATCAC
Db	3241	TCAGGGCTGAGCAAAAGTCTCAGGAGGAAGACATGACCTTGTGACGATGACATCAC
Qy	3301	ACTGTTGCTTATCAAGATGTGCCACAAAATGMAAAAAATTTCTACATCTTACGAGGAAGA

3301	ACTGGTTGCTTATCAAGTGTGCCACAATAATGAAAAAATTTCTTACATCTTACGAGGAAGA	3360
QY	CGAAGCAAGCTACCGCTATTGTGTTTACTTGGAGTAATAGGAGCTGAAATTTGGTCTGAA	3420
DB	CGAAGCAAGCTACCGCTATTGTGTTTACTTGGAGTAATAGGAGCTGAAATTTGGTCTGAA	3420
QY	ATTGAACCTCTAAACTATATGACGAGACCTCGAAGCTCTAGGCCAAATTTCTGAGGGATTC	3480
DB	ATTGAACCTCTCTAAACTATATGACGAGACCTCGAAGCTCTAGGCCAAATTTCTGAGGGATTC	3480
QY	GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATATTTGCAAGGCACCTGACG	3540
DB	GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATATTTGCAAGGCACCTGACG	3540
QY	TTTCTTCTGCTACAGCCTCCAAAGCCCCAAACTTCTCCACAGACATATCCGAAGAACA	3600
DB	TTTCTTCTGCTACAGCCTCCAAAGCCCCAAACTTCTCCACAGACATATCCGAAGAACA	3600
QY	GCCATTGATCTGATTTGGAAGTGGGTTCACTGTTTGGAGCCTTACATGGATGTGTCCGCT	3660
DB	GCCATTGATCTGATTTGGAAGTGGGTTCACTGTTTGGAGCCTTACATGGATGTGTCCGCT	3660
QY	GTTCGATGGGGCTTCTCGAACTTTGTGCCGATGCGCGAGAAACAATTTGCCAAATCATCA	3720
DB	GTTCGATGGGGCTTCTCGAACTTTGTGCCGATGCGCGAGAAACAATTTGCCAAATCATCA	3720
QY	ATGGGGTTGCTCTGAGCCACGAGCTGACTCGGCCCGCTCTGCGAGGCATGCCCTCTCG	3780
DB	ATGGGGTTGCTCTGAGCCACGAGCTGACTCGGCCCGCTCTGCGAGGCATGCCCTCTCG	3780
QY	CTCATTTGCCACCGCAGACCAACCGGCTTCATCACCACTATGCCAATAGCCAAAGAGGTACA	3840
DB	CTCATTTGCCACCGCAGACCAACCGGCTTCATCACCACTATGCCAATAGCCAAAGAGGTACA	3840
QY	CATACGGCTCTTGCAGCAAAATACCCAAATCACAGCAGAAATATGCAACAACAATCTTGCA	3900
DB	CATACGGCTCTTGCAGCAAAATACCCAAATCACAGCAGAAATATGCAACAACAATCTTGCA	3900
QY	CGAGCTAAAGGGGAAATTTTGAGAGTCATTGAAATTTCTTATTTGAAAAGATGCCACAGAT	3960
DB	CGAGCTAAAGGGGAAATTTTGAGAGTCATTGAAATTTCTTATTTGAAAAGATGCCACAGAT	3960
QY	GTGTGCTGATCTCTCGTGAGGTTATGGACATCATTTATGTACTGCTTTGAGGATCTTTTA	4020
DB	GTGTGCTGATCTCTCGTGAGGTTATGGACATCATTTATGTACTGCTTTGAGGATCTTTTA	4020
QY	GTTTAAAAAGAAAGGCTCTTCAAGAAATGTTTCCAGCGCATCTGCAGGTCTTACATGTGACG	4080
DB	GTTTAAAAAGAAAGGCTCTTCAAGAAATGTTTCCAGCGCATCTGCAGGTCTTACATGTGACG	4080
QY	TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGTTCGCCATATGTTTCACTGAGCCCTGAC	4140
DB	TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGTTCGCCATATGTTTCACTGAGCCCTGAC	4140
QY	GACATCCGAGCTGGAATAATGTCAGACATTCATGACACAGGGGACCAATCACTGCACTG	4200
DB	GACATCCGAGCTGGAATAATGTCAGACATTCATGACACAGGGGACCAATCACTGCACTG	4200
QY	GCTTTTGTCTCTGATGGAAAGATATCTTGCCACTACTCAAAACATCTGACAGCCACATTTCT	4260
DB	GCTTTTGTCTCTGATGGAAAGATATCTTGCCACTACTCAAAACATCTGACAGCCACATTTCT	4260
QY	TTTTTGGCAGATGAACACGTCACTGCTGGGAAGCATCGGGAATGCTGAACTCGGCGACCTCAG	4320
DB	TTTTTGGCAGATGAACACGTCACTGCTGGGAAGCATCGGGAATGCTGAACTCGGCGACCTCAG	4320
QY	CTGCGCTGCAATTAACACCTTACCAGGTGCCCTCTGTCAGCCGCGGTCCCCGGGTCCAC	4380
DB	CTGCGCTGCAATTAACACCTTACCAGGTGCCCTCTGTCAGCCGCGGTCCCCGGGTCCAC	4380
QY	AATGCCCTCAAGCTGGCCGGCTCATCTGGACTTCCAAACCGCAACGTCATCTCATGGCC	4440
DB	AATGCCCTCAAGCTGGCCGGCTCATCTGGACTTCCAAACCGCAACGTCATCTCATGGCC	4440

QY 4441 CATGACGGGAGGACGACCGCTTCATGGTCTAA 4473
Db 4441 CATGACGGGAGGAGGACCGCTTCATGGTCTAA 4473

RESULT 2

US-10-093-463-153

Sequence 153, Application US/10093463

Publication No. US20030208039A1

GENERAL INFORMATION:

APPLICANT: Padigar, Muralidhara
APPLICANT: Shenoy, Suresh
APPLICANT: Kekuda, Ramesh
APPLICANT: Gusev, Vladimir
APPLICANT: Pochart, Pascal
APPLICANT: Zhong, Mei
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter
APPLICANT: Smithson, Glenda
APPLICANT: Guo, Xiaojia
APPLICANT: Gerlach, Valerie
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Zerhusen, Bryan
APPLICANT: Tchernev, Velizar
APPLICANT: Gangolli, Bsha
APPLICANT: Vernet, Corine
APPLICANT: Pena, Carol
APPLICANT: Burgees, Catherine
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly
APPLICANT: Gorman, Linda
APPLICANT: Spaderna, Steven
APPLICANT: Voss, Edward
APPLICANT: Malyankar, Uriel
APPLICANT: Anderson, David
APPLICANT: Patturajan, Meera
APPLICANT: Miller, Charles
APPLICANT: Taupier, Raymond J. Jr.
TITLE OF INVENTION: No. US20030208039A1e1 Antibodies that Bind to Antigenic Polypeptides
TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
FILE REFERENCE: 21402-290A (Cura 590AT)
CURRENT APPLICATION NUMBER: US/10/093,463
PRIOR FILING DATE: 2002-06-24
PRIOR APPLICATION NUMBER: 60/283,675
PRIOR FILING DATE: 2001-04-14
PRIOR APPLICATION NUMBER: 60/338,092
PRIOR FILING DATE: 2001-12-03
PRIOR APPLICATION NUMBER: 60/274,281
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/274,101
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/325,681
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/304,354
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/279,995
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 60/294,899
PRIOR FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/287,424
PRIOR FILING DATE: 2001-04-30
PRIOR APPLICATION NUMBER: 60/299,027
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: 60/309,198
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/281,194
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/274,194
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/274,849

PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/330,380
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/275,235
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: 60/288,342
PRIOR FILING DATE: 2001-05-03
PRIOR APPLICATION NUMBER: 60/275,578
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 370
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 153
LENGTH: 4517
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (20)..(4490)
US-10-093-463-153

Query Match 99.7%; Score 4460.2; DB 17; Length 4517;

Best Local Similarity 99.8%; Pred. No. 0; Mismatches 8; Indels 0; Gaps 0;
Matches 4465; Conservative 0;

QY	1	ATGGCAGGAAACAGCGCTTGTCTACCCCATTTGTTTGGGGTCGAAAGCGCCACACAT	60
Db	20	ATGGCAGGAAACAGCGCTTGTCTACCCCATTTGTTTGGGGTCGAAAGCGCCACACAT	79
QY	61	TGCATCTCAGCGGTACTTTTAAACAGATGAGGGGCGACGATCGTAACAGGATGTCCAGAC	120
Db	80	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCGACGATCGTAACAGGATGTCCAGAC	139
QY	121	GCACAAATATGTCTCTGGGATCTTTCAGTAGAATCGCAAAATTAATCTCGACACTGTTG	180
Db	140	GCACAAATATGTCTCTGGGATCTTTCAGTAGAATCGCAAAATTAATCTCGACACTGTTG	199
QY	181	TTTGGTCATACAGCATCAATCACTTGTCTTAAAGCTTGTGCTTCCAGTCAACAAACAG	240
Db	200	TTTGGTCATACAGCATCAATCACTTGTCTTAAAGCTTGTGCTTCCAGTCAACAAACAG	259
QY	241	TATATTGAGTGCAATCTGAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTATGGCAGA	300
Db	260	TATATTGAGTGCAATCTGAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTATGGCAGA	319
QY	301	TCTATTGAATTTACAAAATTTAGCTTGGCACACATCTGGCATACAGTTCTTACAGTTCTCT	360
Db	320	TCTATTGAATTTACAAAATTTAGCTTGGCACACATCTGGCATACAGTTCTTACAGTTCTCT	379
QY	361	GTGGGAATCAGCGAAGGAAAGGCTTTTATGCCACGGACATTACCTGAAATCCTTGT	420
Db	380	GTGGGAATCAGCGAAGGAAAGGCTTTTATGCCACGGACATTACCTGAAATCCTTGT	439
QY	421	GTGGATGTACAGCGCTTGAAGTATATATCTCTTAGTATCAAGATATCACAGACTGG	480
Db	440	GTGGATGTACAGCGCTTGAAGTATATATCTCTTAGTATCAAGATATCACAGACTGG	499
QY	481	ATTAGCTCCATGATATTATTCGATCCCAACCGACACAGAGAGACACAGTGTAGCACTC	540
Db	500	ATTAGCTCCATGATATTATTCGATCCCAACCGACACAGAGAGACACAGTGTAGCACTC	559
QY	541	TGGGTGACTGGCATCTCGAAGGTCTGGATTGTGTACCTCGGAAATAGTGACATGCAGGAT	600
Db	560	TGGGTGACTGGCATCTCGAAGGTCTGGATTGTGTACCTCGGAAATAGTGACATGCAGGAT	619
QY	601	ACTGAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAAGATTGCCAAGCATC	660
Db	620	ACTGAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAAGATTGCCAAGCATC	679
QY	661	TCTTTTGTGCAATTTACAAAGGTCACTTTGTGTTGTTTCCAAATATTGGAGGGTG	720
Db	680	TCTTTTGTGCAATTTACAAAGGTCACTTTGTGTTGTTTCCAAATATTGGAGGGTA	739
QY	721	TTTCGATGCCGGAGACTATTTCCTTGTGTTTTCAGGTCCTAGTGAATAATGCACAGCATGG	780

QY	2941	AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCGACACCTACTTGGGA	3000
Db	2960	AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCGACACCTACTTGGGA	3019
QY	3001	TTGGATAAATTTAGGGCTCCCTTCTGGAGATGCTGGCCGGAAGATGGCAAGATCGATGC	3060
Db	3020	TTGGATAAATTTAGGGCTCCCTTCTGGAGATGCTGGCCGGAAGATGGCAAGATCGATGC	3079
QY	3061	TTGGAGTGAGAGAAGCCGACAGAGCCCTGCTTCTGGCCGAACCTGAGAGAATTTAGCAG	3120
Db	3080	TTGGAGTGAGAGAAGCCGACAGAGCCCTGCTTCTGGCCGAACCTGAGAGAATTTAGCAG	3139
QY	3121	GCAGGAGGAAGAACCAATGATGCTGGGCTTCTTACTTACTCCTAGTACATAGACCAC	3180
Db	3140	GCAGGAGGAAGAACCAATGATGCTGGGCTTCTTACTTACTCCTAGTACATAGACCAC	3199
QY	3181	GTCAATATCACTGGAGTCACATCAGAAAGCCGCGACAGATATCAACACGGCTCCTGATGCC	3240
Db	3200	GTCAATATCACTGGAGTCACATCAGAAAGCCGCGACAGATATCAACACGGCTCCTGATGCC	3259
QY	3241	TCAGGGCTGAAGCAAAAGTCCAGGAGGAAGAGATGACCTTGTGACGATGACATCAC	3300
Db	3260	TCAGGGCTGAAGCAAAAGTCCAGGAGGAAGAGATGACCTTGTGACGATGACATCAC	3319
QY	3301	ACTGGTTGCTTATCAGTGTCCACAAATGABAAAATTTCTACATCTTACGAGGAAGA	3360
Db	3320	ACTGGTTGCTTATCAGTGTCCACAAATGABAAAATTTCTACATCTTACGAGGAAGA	3379
QY	3361	CGGAAGCAAGCTACCGCTATGTTTACTTTGGAGTAATAGAGCTGAAATTTGGTGTGAA	3420
Db	3380	CGGAAGCAAGCTACCGCTATGTTTACTTTGGAGTAATAGAGCTGAAATTTGGTGTGAA	3439
QY	3421	ATTGAACCTCCTAAACTATTGAACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGATTC	3480
Db	3440	ATTGAACCTCCTAAACTATTGAACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGATTC	3499
QY	3481	GGGTTGACTAGTGGTGAATCCAACTACTCGTCCGACAGATACCTTGCAGGCACTGACG	3540
Db	3500	GGGTTGACTAGTGGTGAATCCAACTACTCGTCCGACAGATACCTTGCAGGCACTGACG	3559
QY	3541	TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTTCCACACAGACATATCCGAAAGACA	3600
Db	3560	TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTTCCACACAGACATATCCGAAAGACA	3619
QY	3601	GCCATTGATCTGATTTGGAAGTGGGTTCACTGTTTGGAGCCTTACATGATGTGCGCT	3660
Db	3620	GCCATTGATCTGATTTGGAAGTGGGTTCACTGTTTGGAGCCTTACATGATGTGCGCT	3679
QY	3661	GTTCGTATGGGCTTCTCGAACTTTGTGCGATGCCGAGAAACAACTTGCACACATCACA	3720
Db	3680	GTTCGTATGGGCTTCTCGAACTTTGTGCGATGCCGAGAAACAACTTGCACACATCACA	3739
QY	3721	ATGGGGTTGCTCTGAGCCAGACGCTGACTCGGCCGCTCTGCGAGGCACTGCCCTCTCG	3780
Db	3740	ATGGGGTTGCTCTGAGCCAGACGCTGACTCGGCCGCTCTGCGAGGCACTGCCCTCTCG	3799
QY	3781	CTCATTTGCCACCGCAGACCAACCGCTTCTCATCCACATAGCCAAAGAGGTACACAGA	3840
Db	3800	CTCATTTGCCACCGCAGACCAACCGCTTCTCATCCACATAGCCAAAGAGGTACACAGA	3859
QY	3841	CATACGGCTCTTGACGCAAAATACCCATTCACAGGAGATATGACACAACTCTTGCA	3900
Db	3860	CATACGGCTCTTGACGCAAAATACCCATTCACAGGAGATATGACACAACTCTTGCA	3919
QY	3901	CGAGCTAAAGGGGAAATTTTGGAGTCAATGAAATTTCTTATGAAAGATGCCACACAGAT	3960
Db	3920	CGAGCTAAAGGGGAAATTTTGGAGTCAATGAAATTTCTTATGAAAGATGCCACACAGAT	3979
QY	3961	GTTCGTGATCTTCTCGTGGAGGTTATGGAATCATCATATGATGATGATGATGATGATGAT	4020
Db	3980	GTTCGTGATCTTCTCGTGGAGGTTATGGAATCATCATATGATGATGATGATGATGATGAT	4039
QY	4021	GTAAAAAGAAAGGCTTTCAAGAAATGTTTCCAGGCATCTGAGGTTTCAATGTTCAAGC	4080
Db	4040	GTAAAAAGAAAGGCTTTCAAGAAATGTTTCCAGGCATCTGAGGTTTCAATGTTCAAGC	4099
QY	4081	TATTATGAGCGGAATACAGAAATAGCAATAGCAATAGCAATAGCAATAGCAATAGCAAT	4140
Db	4100	TATTATGAGCGGAATACAGAAATAGCAATAGCAATAGCAATAGCAATAGCAATAGCAAT	4159
QY	4141	GACATCCGACCTGGAAATGTCAGACAATCCATGAGACACAAGGGACCAATCACTGCAGTG	4200
Db	4160	GACATCCGACCTGGAAATGTCAGACAATCCATGAGACACAAGGGACCAATCACTGCAGTG	4219
QY	4201	GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAAACACTGACAGCAATTTCT	4260
Db	4220	GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAAACACTGACAGCAATTTCT	4279
QY	4261	TTTTGGCAGATGAACACGTCACCTGCTGGGAAGCATCGGATCTGAACTCGGACCTCAG	4320
Db	4280	TTTTGGCAGATGAACACGTCACCTGCTGGGAAGCATCGGATCTGAACTCGGACCTCAG	4339
QY	4321	CTGGCTGCTATTAACACCTTACCAGGTGCCCTGTGCGCCGCTGCCCGCTCCAC	4380
Db	4340	CTGGCTGCTATTAACACCTTACCAGGTGCCCTGTGCGCCGCTGCCCGCTCCAC	4399
QY	4381	AATGCCCTCAAGCTGCGCCGCTCATCTGGAATTTCCAAACCGCAACGTCATCTCATGGCC	4440
Db	4400	AATGCCCTCAAGCTGCGCCGCTCATCTGGAATTTCCAAACCGCAACGTCATCTCATGGCC	4459
QY	4441	CATGACGGGAAGAGACCGCTTCTCATGGTCTAA	4473
Db	4460	CATGACGGGAAGAGACCGCTTCTCATGGTCTAA	4492
RESULT 3			
US-10-120-988-91			
; Sequence 91, Application US/10120988			
; Publication No. US20030219745A1			
; GENERAL INFORMATION:			
; APPLICANT: Tang, Y. Tom			
; APPLICANT: Goodrich, Ryle			
; APPLICANT: Liu, Chenghua			
; APPLICANT: Ren, Feiyun			
; APPLICANT: Wang, Dunrui			
; APPLICANT: Drmanac, Radoje T.			
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and			
; TITLE OF INVENTION: Polypeptides			
; FILE REFERENCE: 802CON			
; CURRENT APPLICATION NUMBER: US/10/120,988			
; CURRENT FILING DATE: 2002-04-11			
; PRIOR APPLICATION NUMBER: 09/774,528			
; PRIOR FILING DATE: 2001-01-30			
; NUMBER OF SEQ ID NOS: 441			
; SOFTWARE: pt_FL_genes Version 2.0			
; SEQ ID NO 91			
; LENGTH: 6224			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)..(3597)			
US-10-120-988-91			
Query Match 78.2%; Score 3497; DB 17; Length 6224;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 3508; Conservative 0; Mismatches 0; Indels 1; Gaps 1;			
QY	966	GTTCGTAATTTGTCCTCTGTTACTCGGTTCTTTATGGATGAGAGATATTTCCATAA	1025
Db	117	GTTCGTAATTTGTCCTCTGTTACTCGGTTCTTTATGGATGAGAGATATTTCCATAA	176
QY	1026	ACTGTTAAATTCAGGGTGATTTCTTCGAAAGGTTGAATATTTGGAACATATCAGACAGC	1085
Db	177	ACTGTTAAATTCAGGGTGATTTCTTCGAAAGGTTGAATATTTGGAACATATCAGACAGC	236

Db 2397 GCTGAAAGCTTCCAGGAGGAGCATGACCTTGTGTGACGATGACATCCACTGG 2456
QY 3306 TTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTCATCTTACGAGAAAGCGAA 3365
Db 2457 TTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTCATCTTACGAGAAAGCGAA 2516
QY 3366 GCAAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAATTTGGTCTGGAATTGA 3425
Db 2517 GCAAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAATTTGGTCTGGAATTGA 2576
QY 3426 ACCTCTTAAACTATTGACAGAGCTCGAAGCTCTAGCCAAATTCCTGAGGGAATTCGGGTT 3485
Db 2577 ACCTCTTAAACTATTGACAGAGCTCGAAGCTCTAGCCAAATTCCTGAGGGAATTCGGGTT 2636
QY 3486 GACTAGTGTGGATCAACTACTCTGCTGGCCAGACATATTGCAAGGCACTGACGTTTCT 3545
Db 2637 GACTAGTGTGGATCAACTACTCTGCTGGCCAGACATATTGCAAGGCACTGACGTTTCT 2696
QY 3546 TCTGCTACAGCTTCCGAAGCCCAAACTTCTTCCACACAGCATATCCGAAGACACCCAT 3605
Db 2697 TCTGCTACAGCTTCCGAAGCCCAAACTTCTTCCACACAGCATATCCGAAGACACCCAT 2756
QY 3606 TGATCTGATTGGACGTGGGTTCACTGTTTGGGAGCCTTACATGAGTGTGTCGCTGTTCT 3665
Db 2757 TGATCTGATTGGACGTGGGTTCACTGTTTGGGAGCCTTACATGAGTGTGTCGCTGTTCT 2816
QY 3666 GATGGGGCTTCTCGAACTTTGTGCGATGCCGAGAAACAACTTGGCCAAATCAATGCGG 3725
Db 2817 GATGGGGCTTCTCGAACTTTGTGCCGATGCCGAGAAACAACTTGGCCAAATCAATGCGG 2876
QY 3726 GTTGCCCTCTGAGCCAGCAGCTGACTCGGCCGCTCTGCGAGGCATGCCCTCTCGCTCAT 3785
Db 2877 GTTGCCCTCTGAGCCAGCAGCTGACTCGGCCGCTCTGCGAGGCATGCCCTCTCGCTCAT 2936
QY 3786 TGCCACGCCAGACACCCGCTTCTATCACCACATAGCCAAAGAGGTACACAGACATAC 3845
Db 2937 TGCCACGCCAGACACCCGCTTCTATCACCACATAGCCAAAGAGGTACACAGACATAC 2996
QY 3846 GGTCTTTGACGCAAAATACCAATCAACAGAGAAATATGCAACAACTCTTGGCAGAGC 3905
Db 2997 GGTCTTTGACGCAAAATACCAATCAACAGAGAAATATGCAACAACTCTTGGCAGAGC 3056
QY 3906 TAAAGGGGAATTTTCAGAGTCAATGAAATTTCTTATTGAAAGATGCCACAGATCTCT 3965
Db 3057 TAAAGGGGAATTTTCAGAGTCAATGAAATTTCTTATTGAAAGATGCCACAGATCTCT 3116
QY 3966 GGATCTTCTGAGGTTATGACATCATATTATGTACTGCTTTGAAAGGATCTTTAGTTAA 4025
Db 3117 GGATCTTCTGAGGTTATGACATCATATTATGTACTGCTTTGAAAGGATCTTTAGTTAA 3176
QY 4026 AAGAAAGGCTTTCAAGAAATTTTCCAGCAATCTCGAGGTTCTACATGCTCAGCTATTAA 4085
Db 3177 AAGAAAGGCTTTCAAGAAATTTTCCAGCAATCTCGAGGTTCTACATGCTCAGCTATTAA 3236
QY 4086 TGAGCGGAATCAAGATAGCAGTTGGAGTCCGCAATGTTTCAAGTGGCCCTGACACAT 4145
Db 3237 TGAGCGGAATCAAGATAGCAGTTGGAGTCCGCAATGTTTCAAGTGGCCCTGACACAT 3296
QY 4146 CCGGATCGGAAAATGTACAGCAATCCAATGACACAAAGGACCAATCACTGCACTGGCTTT 4205
Db 3297 CCGGATCGGAAAATGTACAGCAATCCAATGACACAAAGGACCAATCACTGCACTGGCTTT 3356
QY 4206 TGCTCTGTATGGAAGATATCTTGCCACCTACTCAAACTGACAGCAGCAATTTCTTTTGG 4265
Db 3357 TGCTCTGTATGGAAGATATCTTGCCACCTACTCAAACTGACAGCAGCAATTTCTTTTGG 3416
QY 4266 GCAAGTGAACACTCTCTGCGAGCATCGGCATGCTGAATCGGCACCTCAGCTGGG 4325
Db 3417 GCAAGTGAACACTCTCTGCGAGCATCGGCATGCTGAATCGGCACCTCAGCTGGG 3476
QY 4326 CTGCATTTAAACCTTACAGAGTGCCTCTGTGAGCCCGGCTCCCGGGTCCCACAATGC 4385

Db 3477 CTGCATTTAAACCTTACAGAGTGCCTCTGTGAGCCCGGCTCCCGGGCTCCCACAATGC 3536
QY 4386 CTTCAAGCTTGGCCGGCTCATCTGGATTTCCTCAACCGCAACGCTCATCTCATGGCCCATG 4444
Db 3537 CTTCAAGCTTGGCCGGCTCATCTGGATTTCCTCAACCGCAACGCTCATCTCATGGCCCATG 3596
QY 4445 ACGGGAAGGAGCACCCTTTCATGGTCTAA 4473
Db 3597 ACGGGAAGGAGCACCCTTTCATGGTCTAA 3625

RESULT 4

US-10-094-749-1317
; Sequence 1317, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1317
; LENGTH: 3316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1317

Query Match 62.8%; Score 2807.4; DB 17; Length 3316;

Best Local Similarity 92.7%; Pred. No. 0;

Matches 3080; Conservative 0; Mismatches 1; Indels 242; Gaps 3;

QY 1 ATGGCAGGAAACAGCGCTTGTCTTACCCATTGTTCTTTGGGTCGAAAAGCGCCACACAT 60
Db 236 ATGGCAGGAAACAGCGCTTGTCTTACCCATTGTTCTTTGGGTCGAAAAGCGCCACAT 295
QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCCACGATCGTAAACAGATGTACGAC 120
Db 296 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCCACGATCGTAAACAGATGTACGAC 355
QY 121 GGACAAATATGCTCTGGGATCTTTTACAGTAGAACCTGCAAAATTAATCCTCGAGCATTGTG 180
Db 356 GGACAAATATGCTCTGGGATCTTTTACAGTAGAACCTGCAAAATTAATCCTCGAGCATTGTG 415
QY 181 TTGGTCATACAGCATCAATCACTTGTGTTGTCTTAAAGCTTGTGCTTCCAGTGCAAAACAG 240
Db 416 TTGGTCATACAGCATCAATCACTTGTGTTGTCTTAAAGCTTGTGCTTCCAGTGCAAAACAG 475
QY 241 TATATTGTGAGTGATCTGAAAAGTGAGAGATGTGCTCTGGGATGTGATGTGTCAGA 300
Db 476 TATATTGTGAGTGATCTGAAAAGTGAGAGATGTGCTCTGGGATGTGATGTGTCAGA 535

Db	2553	CTGGATCGCTTGGAAATGCTGAACCCCACTGACCGTATCGTTGGCCCTCTGTCAAGA	2612	
Qy	2521	GGAGGCCATATATGCACTGATGCTGCGGGTTAATACAGCTGCTTTGTAACATGTCACAT	2580	
Db	2613	GGAGGCCATATATGCACTGATGCTGCGGGTTAATACAGCTGCTTTGTAACATGTCACAT	2672	
Qy	2581	GGGAAACAGAGTAGGAGGAGCTGCCAGCTCTGAGGAGTAGGAAAGGAACTTAC	2640	
Db	2673	GGGAAACAGAGTAGGAGGAGCTGCCAGCTCTGAGGAGTAGGAAAGGAACTTAC	2732	
Qy	2641	GGAGTCTCCCGTCCCTGACCAACAGCATCTCTCTGTCATCATTTCTTTGGCAATACT	2700	
Db	2733	GGAGTCTCCCGTCCCTGACCAACAGCATCTCTCTGTCATCATTTCTTTGGCAATACT	2792	
Qy	2701	TTAATGAGTATGACCAATGCAATTTTATTTGGTGATCATATGAAGAGGCTCTACAGG	2760	
Db	2793	TTAATGAGTATGACCAATGCAATTTTATTTGGTGATCATATGAAGAGGCTCTACAGG	2852	
Qy	2761	CCACCTAGACCAAGCAACCCAGACCTTTCTAAGGCAAGGGTTCCCTCCACTCCAGT	2820	
Db	2853	CCACCTAGACCAAGCAACCCAGACCTTTCTAAGGCAAGGGTTCCCTCCACTCCAGT	2912	
Qy	2821	AATATTGTGCAAGCAGACGATTTAAACAAAGTTGCTGCACCTGCTCGTCTCGTCTGAT	2880	
Db	2913	AATATTGTGCAAGCAGACGATTTAAAC-----	2937	
Qy	2881	GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTGCTTTTACATACCTGTTCTTAGTA	2940	
Db	2938	-----	2937	
Qy	2941	AATGAAGTTGGAGTCAGTTAGTCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGGGA	3000	
Db	2938	-----AAGGTTGGAGTCAGTTAGTCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGGGA	2993	
Qy	3001	TTGGATAAATTTAGGCTCCCTCTCTGGAGATGCTGGCCGGAAGATCGATGC	3060	
Db	2994	TTGGATAAATTTAGGCTCCCTCTCTGGAGATGCTGGCCGGAAGATCGATGC	3053	
Qy	3061	TTGGAGGTGAGAGCCGACAGGCCCTGCTTTCTGGCGGAACTGAGAGAAATTGAGCAG	3120	
Db	3054	TTGGAGGTGAGAGCCGACAGGCCCTGCTTTCTGGCGGAACTGAGAGAAATTGAGCAG	3113	
Qy	3121	GCAGGCAGGAGAGCCATGATGCTGGGCTCTTACTTACTAGTACATAGACCCAC	3180	
Db	3114	GCAGGCAGGAGAGCCATGATGCTGGGCTCTTACTTACTAGTACATAGACCCAC	3173	
Qy	3181	GTCAATATCACTGGAGTCACATCAGAAAGCCGCGCAGACTATCACCGGCTCCTGATGCC	3240	
Db	3174	GTCAATATCACTGGAGTCACATCAGAAAGCCGCGCAGACTATCACCGGCTCCTGATGCC	3233	
Qy	3241	TCAGGGCTGAAGCAAAAGTCCAGGAGGAGAGAGATGACCTTGTGACGATGACATCACC	3300	
Db	3234	TCAGGGCTGAAGCAAAAGTCCAGGAGGAGAGAGATGACCTTGTGACGATGACATCACC	3293	
Qy	3301	ACTGGTGTCTATCAAGTGTCCC	3323	
Db	3294	ACTGGTGTCTATCAAGTGTCCC	3316	

RESULT 5

US-10-108-260A-1051

; Sequence 1051, Application US/10108260A

; Publication No. US20040005560A1

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: HI-A0106

; FILE REFERENCE: US 20040005560A1el full length cdna

; CURRENT APPLICATION NUMBER: US/10/108,260A

; CURRENT FILING DATE: 2002-03-27

; NUMBER OF SEQ ID NOS: 5458

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1051

;	LENGTH: 2578
;	TYPE: DNA
;	ORGANISM: Homo sapiens
;	US-10-108-260A-1051
Query Match	10.6%; Score 474.6; DB 17; Length 2578;
Best Local Similarity	55.1%; Pred. No. 5.1e-132;
Matches	999; Conservative 0; Mismatches 799; Indels 15; Gaps 3;
Qy	29 TTGTTCTTTGGGGTCGAAAGCGCCACACATTCGATCTCAGGGTACTTTTAAACAGATG 88
Db	276 TGSCACTCTGGGGACAGAAAGGCCCTCCCCACAGCATCACTGCCATCATGATCACTGATG 335
Qy	89 ATGGGSCCAGCATCGTAAACAGAGTGTACGACGGACAAATATGTCTCTGGGATCTTTTCAAG 148
Db	336 ACAGGGAACGATTTGTGACTTGAAGTCAAGAGGGTCAGCTCTGTCTCTGGAACTCTCTCAC 395
Qy	149 TAGAACTGCAAAATTAATCCTCGAGCACTGTTGTTGGTCATACAGCATCAATCACTTTGTT 208
Db	396 ATGAACTTAAAGATTTTCAGCGAAAGAACTCCTATTTGGTCAATTCAGCTTCGGTAAATGTT 455
Qy	209 TGTCTAAAGCTTTGCTTCCAGTGA CAAACAGATATATTGTGAGTGCATCTGAAAGTGGAG 268
Db	456 TGCGAAGAGCAAGGGACTTCTCTAAACAGCCCTACATTTGTTAGTGTCTGCTGAAAAATGGGG 515
Qy	269 AGATGTGCCCTCTGGGATGTGAGTGATGGCAGATGTATTCGAAATTTACAAAAATTAGCTTGCA 328
Db	516 AGATGTGCTTTGGAAATGTCCAAATGGACAGTGGCTGGAGAGGCTACACTTCCTTAC 575
Qy	329 CACATACCTGGCATACAGTTCTACAGTTCTCTGTTGGGAATCAGCGAGAGGAGGCTTT 388
Db	576 GGCACACTGCAATCTGTTATTACCACTGCTGCTCCGGATGACAGGAGAGGCTGGCTTC 635
Qy	389 TATGCCACCGACATTTACCTGAAATCCTTTGTTGGATGCTACCAGCCTTTGAAGTATTAT 448
Db	636 TTTGTTGTGGAGAATATCAAGATGTCCTTATTAATTGATGCCAAAACTTTTGGCTGTTGTTTC 695
Qy	449 ACTCCTCTTAGTATCAAGATATCAGCAGACTGGATTAGCTCCATGATGATTTATTCGATCCC 508
Db	696 ACAGTTTATGATCATCTCAGTTTCTGACTGGATCACTGCAATGCTGATGTTTCTACTCCA 755
Qy	509 ACCGAACACAGAGGACACAGTGGTAGCACTCTCGGTGATCGGCATCTTGAAGGCTGGGA 568
Db	756 TGAGAAATTCAGAAAGATTCTCTCTGGTGTATCAGTAGCTGGTAGCTCAAAAGTATGGG 815
Qy	569 TTGTTACCTCGGAAATAAGTGACATGACGATCTCAGCCATATTTTGGAGGAGGATCCA 628
Db	816 ATCTTTTCTCTATCTATCAACAGCATTCAGGAAAGCAAGATGTCTATGAAAAAGAAATCCA 875
Qy	629 AACCAATTTATGTGAGAAATTCGCAAGCATCTCTTTTGTGCAATTTACACAAAGGCTCAC 688
Db	876 AGTTTCTTGAGTCTTGAATCTGCCAGCAATTCGATTTTGCATATATCTGAGAGACTTC 935
Qy	689 TTTTGGTGTGTGTTCCAAATATTTGAGGGTGTTCGATGCCGAGACTATTCCTTGTGTTGT 748
Db	936 TATTGGTGTGTTTCTAAATGTTGGAAGGTTTATGATTTATGATTTTTCCTTCTGTCG 995
Qy	749 GTTCAGGTCCTAGTGAATTTGACACACATGACCGGGGGGACTTTGTTCTCATCAGATA 808
Db	996 TGACTGAAGTTAGTAGAAATGGGCAGTTCTTTTGTGCTGGAGAGGATGTTGCTGCTCACA 1055
Qy	809 AAGTCATCATTTTGGACAGAAAAATGGCAAGTTATATTTACAAACTACCTGCCAGTGTCC 868
Db	1056 GAATCTCTCATCTGGACAGAAAGATGGTCA CAGTTACATCTATCAGCTGCTGAAACAGTGGG 1115
Qy	869 TTCCAGCTAGTGATTTCAITTCGAGTGTGGGAGGAGGAGTGTGAAAAATTTAATTCCTC 928
Db	1116 TTTCAAAAAG-----CATATACCTGCTGATGGAAGAGTGTCTTAAAGAGACCATTTATC 1169
Qy	929 CTGTACACATATCTCTTGGATCGAAAGATGAAAGATTGCTTAAATTTGCTCTCTCTGTTA 988
Db	1170 CTCATTTACTGTGCTCTACTTCTGTGCGAAAAAATAAGGAACAGAGCCGCTCTTGTGTTA 1229


```
QY 989 CTGGTCTCTTATGGATGACAGAAATATTTCCATAAACTGTTAAATTCAGGGTGATTCCT 1048
Db 1230 TGGCTACATGAATGAAGAAAGAGCCCTTTTACAAAGTACTTTCTCTGGAGAGTCT 1289
QY 1049 CTGAAGGTTGAATATTTGGAACATATCAGACACAGCTG-----ATAACAGGGAAGTG 1102
Db 1290 CAGGAGAAATATCTTTGTGGCACAATCCTGATGTTCTGTATCCAGTTTGATGGTTCTC 1349
QY 1103 AAGAGGGCTGGCAATGACAACTCTATATAGTTTGGCAAGGCAATTTGATAAATCAATC 1162
Db 1350 CTAGAGAGATACCAGTAAGTACGCCACCTGGACTCTTCAAGATAAATTTTGATAAGCATGATA 1409
QY 1163 CTGTCTCTGCTGGAATATATAGATCAGCTGAGTGTGATTCCTCAATAGTAAGACCTCTTA 1222
Db 1410 CTATGTCACAAAGTATTAATGACTATTTCTCTGGGCTTAAAGATGGGGCAGGAACGCTG 1469
QY 1223 AAGTAACCTGCAAGTGTGTACATACACAGACATGACGACTGTTTGTGGTGTGGAAGATG 1282
Db 1470 TAGTCACTTCATCAGAGTATATTTCCAGTCTTGTATAAATTAATATGTGGCTGTGAAGATG 1529
QY 1283 GAAGCATAGTATATGTAACCTGACCAACAGACAGACGCTCGTGTGATCAAGGGAAC 1342
Db 1530 GGACAAATATATCAACAGGCTTGAATGCTGCCAAAGCAAGACTTCTGGAAGGTGGTT 1589
QY 1343 ACATGCTCAGAGAGGTTGGCCACCTCCAGACACTCCGTGTGATCGGGAACAAAGTCA 1402
Db 1590 CTTTAGT---AAAGATTTCTCCCTCATAAAGTCTTAAAGGCCACCAAGGTGTC 1646
QY 1403 CATGTTTGTCTATATCTCATCAGTCTCAGCTCGGTATGATCAAGATACCTCATATCTG 1462
Db 1647 CTTCAATCTACCAACATAGGTCTCTCTCGAAATAGACCAAGTTGGATGTTGTTCTG 1706
QY 1463 GAGGTGTGGATTTTTCAGTCATAAATTTGGGACATATTTCTCGAGAAATGAACATATCT 1522
Db 1707 GGGACCTGGACTCATGTGTGATCTGTGGGATATCTTACTGAAGAAATTTTGCATAAAT 1766
QY 1523 TCTGTGTTATGTTGTGAGATTAATCACTTCTAGTTCACCTCGAAGAACTGATGTCGAA 1582
Db 1767 TCTTTTGGAGCTGGTCCAGTAACAGTCTTTTGTGATGTCAACAGAGAGTTTAAACTAA 1826
QY 1583 GAGTACAGCACTGCATCTGCTGTAGCCAGTACCACTCAGTAGGACTCTTAAGTTGC 1642
Db 1827 GGGGTGACGAGATAATTTGCTGTGTGCGGTGACCATTCGGTGGCTCTCTTCACTTG 1886
QY 1643 GAGAGAAAAATGCAATAATTTGGCATCTCGTCACCTTTTCTCTATTTCAAGTAATCAAT 1702
Db 1887 AGGGAAGAGTTCCTCGCATGCCCGGAAGCACCTTTTCTGTGAGGATGATAAAT 1946
QY 1703 GAGAGGCTTCTGATGATTAACCTGGTGGTGGGTGTTCAAGTGTCTGTGACGTCGTC 1762
Db 1947 GGCACCCGGTTGAGAAATTTTAAATTTGTTGGATGTGCAGATGACTCAGTTTATATCTGG 2006
QY 1763 AAATGATACCTGTCATGATGATGATGTTGTGTGATGGGATAACAGCAGTTGAGATCTAA 1822
Db 2007 AAATGGAACAGGCATTTTGAAGAAGATGAGACAGAGAGAGACGACCAAGTAATTTCTTA 2066
QY 1823 ACGCTTGTGTGATGA 1835
Db 2067 ATTTGTTGTGATGA 2079
```

RESULT 6

```
US-10-029-386-7862
; Sequence 7862, Application US/10029386
; Publication No. US200301947041
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
```

```
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7862
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR18.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P22813, EVALUE 2.90e-01
; OTHER INFORMATION: NT HIT: AF188125.1, EVALUE 3.00e-64
; OTHER INFORMATION: EST_HUMAN HIT: AW896708.1, EVALUE 7.00e-20
US-10-029-386-7862
```

```
Query Match 3.4%; Score 150.2; DB 16; Length 585;
Best Local Similarity 74.9%; Pred. No. 6.3e-34;
Matches 188; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
QY 635 TTATTGTGCAAGATTCGCAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACTTTTGG 694
Db 186 TTTTATTTTAACTTGAATGTCATGGAATCTTACTTTTACAAATATGGAGTAGATT 245
QY 695 TTGTGTGTTCCAAATATTTGGAGGGTGTTCGATGCCGAGACTATTCCTTTGTGTTCAG 754
Db 246 GTGAATTTCTTTCTTTCAGAGTGTTCGATGCCGAGACTATTCCTTTGTGTTCAG 305
QY 755 GTCTAGTAGTGAATGACAGACATGACCGGGGGGACTTTGTCTCATCAGATAAAGTCA 814
Db 306 GTCTAGTAGTGAATGACAGACATGACCGGGGGGACTTTGTCTCATCAGATAAAGTCA 365
QY 815 TCATTGGACAGAAATGGCAAGTTATATTACAACTACTCTGCCAGTTGCCCTTCCAG 874
Db 366 TCATTGGACAGAAATGGCAAGTTATATTACAACTACTCTGCCAGTTATGCAGCAA 425
QY 875 CTAGTGATTTCA 885
Db 426 GTATATAATTA 436
```

RESULT 7

```
US-10-029-386-21562
; Sequence 21562, Application US/10029386
; Publication No. US200301947041
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21562
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR18.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P22813, EVALUE 3.00e-01
; OTHER INFORMATION: NT HIT: AF188125.1, EVALUE 5.00e-64
; OTHER INFORMATION: EST_HUMAN HIT: AW896708.1, EVALUE 2.00e-20
US-10-029-386-21562
```

```
Query Match          3.3%; Score 147.2; DB 16; Length 227;
Best Local Similarity 92.3%; Pred. No. 2.6e-33; Indels 0; Gaps 0;
Matches 155; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 718 GTGTTGATGCGGAGACATTCTCTTGTGTGTTTCAGGTCCTAGTGAATAATGGACAGACA 777
    |||||||
Db 1 GTGTTGATGCGGAGACATTCTCTTGTGTGTTTCAGGTCCTAGTGAATAATGGACAGACA 60

QY 778 TGAACGGGGGGGACATTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAA 837
    |||||||
Db 61 TGAACGGGGGGGACATTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAA 120

QY 838 AGTTATATTACAAACTACTGCCAGTTCCTTCCAGCTAGTGATTCA 885
    |||||||
Db 121 AGTTATATTACAAACTACTGCCAGTATGCAGCAAGTAATAAATTA 168
```

```
RESULT 8
US-10-027-632-96119
; Sequence 96119, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 96119
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-96119
```

```
Query Match          2.8%; Score 125.4; DB 13; Length 585;
Best Local Similarity 99.2%; Pred. No. 2.1e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCAAATTTTGGAGGAGGAATCCAAACCAATTTTGTGTCAGAAATGGC 652
    |||||||
Db 342 TTGAGGATCTGAGCCAAATTTTGGAGGAGGAATCCAAACCAATTTTGTGTCAGAAATGGC 401

QY 653 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTTCCAAATATT 712
    |||||||
Db 402 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTTCCAAATATT 461

QY 713 GGAGGGT 719
    |||||||
Db 462 GGAGGGT 468
```

```
RESULT 9
US-10-027-632-96119
; Sequence 96119, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 96119
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-96119

Query Match          2.8%; Score 125.4; DB 17; Length 585;
Best Local Similarity 99.2%; Pred. No. 2.1e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCAAATTTTGGAGGAGGAATCCAAACCAATTTTGTGTCAGAAATGGC 652
    |||||||
Db 342 TTGAGGATCTGAGCCAAATTTTGGAGGAGGAATCCAAACCAATTTTGTGTCAGAAATGGC 401

QY 653 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTTCCAAATATT 712
    |||||||
Db 402 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTTCCAAATATT 461

QY 713 GGAGGGT 719
    |||||||
Db 462 GGAGGGT 468

RESULT 10
US-10-027-632-320926
; Sequence 320926, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 320926
; LENGTH: 590
```

```
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-320926

Query Match          2.8%; Score 125.4; DB 13; Length 590;
Best Local Similarity 99.2%; Pred. No. 2.1e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 652
      |||||
Db 346 TTCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 405

QY 653 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
      |||||
Db 406 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 465

QY 713 GGAGGGT 719
      |||||
Db 466 GGAGGGT 472

RESULT 11
US-10-027-632-320926
; Sequence 320926, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 320926
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-320926

Query Match          2.8%; Score 125.4; DB 17; Length 590;
Best Local Similarity 99.2%; Pred. No. 2.1e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 652
      |||||
Db 346 TTCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 405

QY 653 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
      |||||
Db 406 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 465

QY 713 GGAGGGT 719
      |||||
Db 466 GGAGGGT 472

RESULT 12
US-10-027-632-237407
```

```
; Sequence 237407, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237407
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-237407

Query Match          2.8%; Score 125.4; DB 13; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 652
      |||||
Db 241 TTCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 300

QY 653 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
      |||||
Db 301 AAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
      |||||
Db 361 GGAGGGT 367

RESULT 13
US-10-027-632-237408
; Sequence 237408, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237408
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237408

Query Match      2.8%; Score 125.4; DB 13; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATACGTAGCCCAATTTGAGGAGGAATCCAAACCAATTTATTGTGCAAAATTGCC 652
Db 241 TTCCAGGATACGTAGCCCAATTTGAGGAGGAATCCAAACCAATTTATTGTGCAAAATTGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
Db 361 GGAGGGT 367

RESULT 14
US-10-027-632-237407
; Sequence 237407, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237407
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237407

Query Match      2.8%; Score 125.4; DB 17; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATACGTAGCCCAATTTGAGGAGGAATCCAAACCAATTTATTGTGCAAAATTGCC 652
Db 241 TTCCAGGATACGTAGCCCAATTTGAGGAGGAATCCAAACCAATTTATTGTGCAAAATTGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
Db 361 GGAGGGT 367

RESULT 15
US-10-027-632-237408
; Sequence 237408, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237408
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237408

Query Match      2.8%; Score 125.4; DB 17; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATACGTAGCCCAATTTGAGGAGGAATCCAAACCAATTTATTGTGCAAAATTGCC 652
Db 241 TTCCAGGATACGTAGCCCAATTTGAGGAGGAATCCAAACCAATTTATTGTGCAAAATTGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
Db 361 GGAGGGT 367

Search completed: July 2, 2005, 13:41:56
Job time : 2395 secs
```


QY	1146	ATTGTGATAAACTGAATCCTTGTCTCTGCTGGAATTATAGATCAGCTGAGTGTGATTCGCCAA	1205
Db	297	ATTGTGATAAACTGAATCCTTGTCTCTGCTGGAATTATAGATCAGCTGAGTGTGATTCGCCAA	356
QY	1206	TAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTATACATACAGACGACATGGACGACTTGT	1265
Db	357	TAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTATACATACAGACGACATGGACGACTTGT	416
QY	1266	TTGTGTCGTGAAGATGGAAGCATAGTATTATGACCTGCGCACACAGACGCGCCATAGTACA	1325
Db	417	TTGTGTCGTGAAGATGGAAGCATAGTATTATGACCTGCGCACACAGACGCGCCATAGTACA	476
QY	1326	GCTGTTGCAAGGGGAAACATATGCTCAGAGAGGTGTGGCCACCTCACAAGAACATCCGTGG	1385
Db	477	GCTGTTGCAAGGGGAAACATATGCTCAGAGAGGTGTGGCCACCTCACAAGAACATCCGTGG	536
QY	1386	TCATCGGAACAAAGTACATGTTTGTATATCTCATCAGGTCTCAGCTCGGTATGATCA	1445
Db	537	TCATCGGAACAAAGTACATGTTTGTATATCTCATCAGGTCTCAGCTCGGTATGATCA	596
QY	1446	AAGATACCTGATATCTGGAAGTGTGATTTTTCAGTCATAATTTGGGACATATTTCTGG	1505
Db	597	AAGATACCTGATATCTGGAAGTGTGATTTTTCAGTCATAATTTGGGACATATTTCTGG	656
QY	1506	AGAAATGAAACATATCTTCTGTGTTTCAATGCTGTGAGATTAATCAACTTCTAGTTCAC	1565
Db	657	AGAAATGAAACATATCTTCTGTGTTTCAATGCTGTGAGATTAATCAACTTCTAGTTCAC	716
QY	1566	TGAAAACCTGAGTGCAGAGTACAGCACTGCATCTGCTGTAGCCAGTGCACACTCAGT	1625
Db	717	TGAAAACCTGAGTGCAGAGTACAGCACTGCATCTGCTGTAGCCAGTGCACACTCAGT	776
QY	1626	AGGACTTCTAAGTTTGCAGAGAAAAATGCATAATGTTGGCATCTCGTCACCTTTTCC	1685
Db	777	AGGACTTCTAAGTTTGCAGAGAAAAATGCATAATGTTGGCATCTCGTCACCTTTTCC	836
QY	1686	TATTCAAGTAACTCAATGGAAGGCTTCTGATGATTAATACCTGGTGGTGGTGTTCAGATGG	1745
Db	837	TATTCAAGTAACTCAATGGAAGGCTTCTGATGATTAATACCTGGTGGTGGTGTTCAGATGG	896
QY	1746	TTCTGTGTACGCTCTGGCAAAATGATATCTGTCATTTGGATGCGTGTGTGATGGGGATAAC	1805
Db	897	TTCTGTGTACGCTCTGGCAAAATGATATCTGTCATTTGGATGCGTGTGTGATGGGGATAAC	956
QY	1806	AGCAGTTGAGATTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTGATTCACCTTAG	1865
Db	957	AGCAGTTGAGATTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTGATTCACCTTAG	1016
QY	1866	TCATCCAGCAGTCAACCTAAACAGCTATGACGAGACGTAGTCTTGCTGCTTAAAAA	1925
Db	1017	TCATCCAGCAGTCAACCTAAACAGCTATGACGAGACGTAGTCTTGCTGCTTAAAAA	1076
QY	1926	TATGGCCATCATAAAGCTACAAACCTTGCACCTAACCTCTGCTGCTTCTGAGGACATCTGA	1985
Db	1077	TATGGCCATCATAAAGCTACAAACCTTGCACCTAACCTCTGCTGCTTCTGAGGACATCTGA	1136
QY	1986	CAAGGGAAATTTTACCTAAATPATTTCTCATAACTCCCTGATGGTTCAAGCAATAAAGCAAA	2045
Db	1137	CAAGGGAAATTTTACCTAAATPATTTCTCATAACTCCCTGATGGTTCAAGCAATAAAGCAAA	1196
QY	2046	CCTAAACGAGCCCGACATACATGCTGATTTCTTTGATGGAGCGTTGATTTTCACT	2105
Db	1197	CCTAAACGAGCCCGACATACATGCTGATTTCTTTGATGGAGCGTTGATTTTCACT	1256
QY	2106	CCTGACTGGAAGAGCCTCTAGGCGGACATCTGCTCTTAATTTCCCGAGAGAAATTTGCAAAA	2165
Db	1257	CCTGACTGGAAGAGCCTCTAGGCGGACATCTGCTCTTAATTTCCCGAGAGAAATTTGCAAAA	1316
QY	2166	AGCATCTGGCAGTTTCAGACAAAAGGGGCTCTTTTTTAACTGGAAAAACGACGACGATCT	2225
Db	1317	AGCATCTGGCAGTTTCAGACAAAAGGGGCTCTTTTTTAACTGGAAAAACGACGACGATCT	1376
QY	2226	CTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGAGAACACCTCTTGTGATGATGA	2285
Db	1377	CTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGAGAACACCTCTTGTGATGATGA	1436
QY	2286	AGAGGAGGATGAGGAGATTAATGAGGCAGAGAAAGGAGAAAGTGATCTCTGAATATCGGTC	2345
Db	1437	AGAGGAGGATGAGGAGATTAATGAGGCAGAGAAAGGAGAAAGTGATCTCTGAATATCGGTC	1496
QY	2346	CAGCAAAATCAAGGCCATTGACCCCTATTAGAAATATAATTTAACTATGGAACACTGCAAGCT	2405
Db	1497	CAGCAAAATCAAGGCCATTGACCCCTATTAGAAATATAATTTAACTATGGAACACTGCAAGCT	1556
QY	2406	GTTTATGCTCCTCTTCAAGCTGCGGTTTGAATGAAGTACTGGAATGAAGTTTGCCTGGA	2465
Db	1557	GTTTATGCTCCTCTTCAAGCTGCGGTTTGAATGAAGTACTGGAATGAAGTTTGCCTGGA	1616
QY	2466	TGCGCTTGGAAATGCTGAAACCCCACTGCAACCGTATCGTTTGGCTCTTCTCAAGAGGAG	2525
Db	1617	TGCGCTTGGAAATGCTGAAACCCCACTGCAACCGTATCGTTTGGCTCTTCTCAAGAGGAG	1676
QY	2526	CCATATGCTCAGTACTGCTGCGGTTTATTAATCAGCTGCTTGTAACTGTCATATGGAA	2585
Db	1677	CCATATGCTCAGTACTGCTGCGGTTTATTAATCAGCTGCTTGTAACTGTCATATGGAA	1736
QY	2586	AACAGAGTAGGAAGGAGCTGCCAGCGTCTCAGGAGTAGGAAAGGGAACCTTACCGEAGT	2645
Db	1737	AACAGAGTAGGAAGGAGCTGCCAGCGTCTCAGGAGTAGGAAAGGGAACCTTACCGEAGT	1796
QY	2646	GTCCGCTGCGGTCAACACACAGCATCTCTGTCTATCATTTTCTTTGGGCAAACTCTTTAAT	2705
Db	1797	GTCCGCTGCGGTCAACACACAGCATCTCTGTCTATCATTTTCTTTGGGCAAACTCTTTAAT	1856
QY	2706	GAGTATGACCAATGCACTTTTATTTGTTGATCATATGAAGAAGGCTCTTACGAGCCACC	2765
Db	1857	GAGTATGACCAATGCACTTTTATTTGTTGATCATATGAAGAAGGCTCTTACGAGCCACC	1916
QY	2766	TAGACCAAGCACCCAGACCTTTCTAAAGCAAGGGTCTCCCTCCAACTTCCAGTAATAT	2825
Db	1917	TAGACCAAGCACCCAGACCTTTCTAAAGCAAGGGTCTCCCTCCAACTTCCAGTAATAT	1976
QY	2826	TGTGCAAGGACAGATTAAACAAAGTTGCTGACCTGTCGTTTCCGCTCGGTCTGATGCTGA	2885
Db	1977	TGTGCAAGGACAGATTAAACAAAGTTGCTGACCTGTCGTTTCCGCTCGGTCTGATGCTGA	2036
QY	2886	TCACTCTGCTCTGACCCCTCTTCTGCTTACATCTGTTTCTTCTTCTTCTTCTTCTTCT	2945
Db	2037	TCACTCTGCTCTGACCCCTCTTCTGCTTACATCTGTTTCTTCTTCTTCTTCTTCTTCT	2096
QY	2946	AGTGTGAGTCAAGTTAGCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGCGGATGGA	3005
Db	2097	AGTGTGAGTCAAGTTAGCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGCGGATGGA	2156
QY	3006	TAAATTTAGGCTTCCCTTCTGAGAGTGTGCGCCGAGAGATGGCAAGATCGATGCTTGA	3065
Db	2157	TAAATTTAGGCTTCCCTTCTGAGAGTGTGCGCCGAGAGATGGCAAGATCGATGCTTGA	2216
QY	3066	GGTGAGAGAGCCGACACAGCCCTCTGCTTCTGCGGAACTGAGAGAAATTTGAGCAGCAGG	3125
Db	2217	GGTGAGAGAGCCGACACAGCCCTCTGCTTCTGCGGAACTGAGAGAAATTTGAGCAGCAGG	2276
QY	3126	CAGGAAGGAGCCATTTGATGCTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	3185
Db	2277	CAGGAAGGAGCCATTTGATGCTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	2336
QY	3186	ATCACCTGGAGTCAATCAAGAGCCGCGAGATATCAACAGGCTCTTGTGATGCTCTCAGG	3245
Db	2337	ATCACCTGGAGTCAATCAAGAGCCGCGAGATATCAACAGGCTCTTGTGATGCTCTCAGG	2396
QY	3246	GCTTGAAGCAAAAGTCCAGGAGGAGCATGACCTTGTGATGAGTATGATGATGATGATGAT	3305
Db	2397	GCTTGAAGCAAAAGTCCAGGAGGAGCATGACCTTGTGATGAGTATGATGATGATGATGAT	2456
QY	3306	TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTTCTACATCTTACGAGGAAAGACGAA	3365

Db 2457 TTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCTTTACGAGGAAGACGGAA 2516
QY 3366 GCAAGCTACCGCTATTGTTTTTACTTGGAGTAATAGGAGCTGAATTTGGTCTCTGAAATTTGA 3425
Db 2517 GCAAGCTACCGCTATTGTTTTTACTTGGAGTAATAGGAGCTGAATTTGGTCTCTGAAATTTGA 2576
QY 3426 ACCTCTTAACCTATTGACGAGACTCGAAGCTCTAGCCTTCTAGCCTTCTGAGGGATTCGGTT 3485
Db 2577 ACCTCTTAACCTATTGACGAGACTCGAAGCTCTAGCCTTCTAGCCTTCTGAGGGATTCGGTT 2636
QY 3486 GACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATACCTTGAAGSCACTGACGTTTCT 3545
Db 2637 GACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATACCTTGAAGSCACTGACGTTTCT 2696
QY 3546 TCTGCTACAGCTTCCAAAGCCCAAACTTCTCCACAGCAGCTATCCGAAGAACAGCCAT 3605
Db 2697 TCTGCTACAGCTTCCAAAGCCCAAACTTCTCCACAGCAGCTATCCGAAGAACAGCCAT 2756
QY 3606 TGATCTGATTGGAGTGGTTCACTGTTGGGAGCCTTACATGATGTGTCGCTGTCT 3665
Db 2757 TGATCTGATTGGAGTGGTTCACTGTTGGGAGCCTTACATGATGTGTCGCTGTCT 2816
QY 3666 GATGGGCTTCTCGAACTTTGCGGATGCGGAGAAACAACTTGCACCAATCAATGGG 3725
Db 2817 GATGGGCTTCTCGAACTTTGCGGATGCGGAGAAACAACTTGCACCAATCAATGGG 2876
QY 3726 GTTGCCCTCTGAGCCAGCAGCTGACTCGGCCCTCTGCGAGGCGATGCCCTCTCGCTCAT 3785
Db 2877 GTTGCCCTCTGAGCCAGCAGCTGACTCGGCCCTCTGCGAGGCGATGCCCTCTCGCTCAT 2936
QY 3786 TGCACCGCCAGACACCCCGCTTCTATCACCATAGCCAAAGAGGTACACAGACATAC 3845
Db 2937 TGCACCGCCAGACACCCCGCTTCTATCACCATAGCCAAAGAGGTACACAGACATAC 2996
QY 3846 GGCTCTGACGAAATACCAATTCAGCAGAGATATGCACAACTCTTGCAGGAC 3905
Db 2997 GGCTCTGACGAAATACCAATTCAGCAGAGATATGCACAACTCTTGCAGGAC 3056
QY 3906 TAAAGGGGAAATTTGAGAGTCAATGAAATTTCTATTGAAAGATGCCACAGATGTTGT 3965
Db 3057 TAAAGGGGAAATTTGAGAGTCAATGAAATTTCTATTGAAAGATGCCACAGATGTTGT 3116
QY 3966 GGATCTCTCGTGAGGATTAGGACATCAATATGTAATGCTGCTTGAAGGATCTTTAGTTAA 4025
Db 3117 GGATCTCTCGTGAGGATTAGGACATCAATATGTAATGCTGCTTGAAGGATCTTTAGTTAA 3176
QY 4026 AAGAAAGTCTTCAAGAAATGTTTCCAGCAGCTGAGGTTCTACATGCTGAGCTATTA 4085
Db 3177 AAGAAAGTCTTCAAGAAATGTTTCCAGCAGCTGAGGTTCTACATGCTGAGCTATTA 3236
QY 4086 TGAGCGGAATCAGAAATAGCAGTGGAGCTCGCCATGTTTCACTGAGGCTGTACACAT 4145
Db 3237 TGAGCGGAATCAGAAATAGCAGTGGAGCTCGCCATGTTTCACTGAGGCTGTACACAT 3296
QY 4146 CCGGACTGGAATAATGTGAGCAATCCATGAGCAAAAGGGACCAATCACTGAGTGCGCTT 4205
Db 3297 CCGGACTGGAATAATGTGAGCAATCCATGAGCAAAAGGGACCAATCACTGAGTGCGCTT 3356
QY 4206 TGCTCTGTATGGAAGATATCTGCGACCTACTCAAACTGACGAGCCACATTTCTTTTGG 4265
Db 3357 TGCTCTGTATGGAAGATATCTGCGACCTACTCAAACTGACGAGCCACATTTCTTTTGG 3416
QY 4266 GCAGATGAACAGCTCACTGCTGGGAAGCATCGGCATGCTGAATCGGCACTCAGCTGGG 4325
Db 3417 GCAGATGAACAGCTCACTGCTGGGAAGCATCGGCATGCTGAATCGGCACTCAGCTGGG 3476
QY 4326 CTGCATTAAACCTACAGGTGCGCTGTGACGCGCGCTGCCCGGCTCCACAAATGC 4385
Db 3477 CTGCATTAAACCTACAGGTGCGCTGTGACGCGCGCTGCCCGGCTCCACAAATGC 3536
QY 4386 CTTCAAGCT-GGCCGCGCTCATCTGACTTCCAAACCGCAACGTCATCTCTCATGGCCATG 4444

Db 3537 CCTCAAGCTGGGCGGCTCATCTGCACTTCCAAACGCAACGTCATCTCTCATGGCCATG 3596
QY 4445 ACGGGAAGGAGCACCCTTTCTATGTTCTAA 4473
Db 3597 ACGGGAAGGAGCACCCTTTCTATGTTCTAA 3625
RESULT 2
US-09-513-999C-14504
; Sequence 14504, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14504
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13
; OTHER INFORMATION: n=a, g, c or t
; NAME/KEY: misc_feature
; LOCATION: 20
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 230
; OTHER INFORMATION: s=g or c
US-09-513-999C-14504
Query Match 2.3%; Score 103; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCAGGAAACAGCCTTCTTACCATTGTTCTTTGGGTGCGAAAGCGCCACACAT 60
Db 254 ATGGCAGGAAACAGCCTTCTTACCATTGTTCTTTGGGTGCGAAAGCGCCACACAT 313
QY 61 TGCATCTCAGCGTACTTTTAAACAGATGATGGGCCACCATCG 103
Db 314 TGCATCTCAGCGTACTTTTAAACAGATGATGGGCCACCATCG 356
RESULT 3
US-08-188-582-17
; Sequence 17, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAPs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2152 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2112
US-08-188-582-17

Query Match 1.1%; Score 48; DB 1; Length 2152;
Best Local Similarity 57.2%; Pred. No. 0.0031;
Matches 127; Conservative 0; Mismatches 90; Indels 5; Gaps 2;
QY 4034 GTCTTCAAGATGTTTCCAGCATCTGCAGGTTCTACATGGTCAGTATTATGACGGGA 4093
DB 1566 GCCTTTAAGATATTTCGCGGCATCTTGCTGATGGAATTGTACCAGATTCCATCCAA 1625
QY 4094 ATCAGAGATAGCAGTTGGAGCTCGCC--ATGGTTTCAAGTGGCCCTGTACGACATCGGAC 4151
DB 1626 TTCTAATTATGTTGCTACGGGCTCTGCAGACAGACTGTGGGCTCTGGAGCTCTGAA 1685
QY 4152 TGGAAATGT---CAGACATCCATGGACACAAAGGACCAATCACTGCACTGCTTTTC 4208
DB 1686 TGGTAATGTTGAAGGATCTTCACTGGACACAGGGACCAATTTCATTCCTTGACATTTTC 1745
QY 4209 TCTGTATGGAAGATATCTTGGCACCCTACTCAAAACACTGACAG 4250
DB 1746 TCCCAATGGAGATCTCTGGCTACAGGACCAACAGATGGCAG 1787

RESULT 4
US-08-646-715-17
Sequence 17, Application US/08646715
Patent No. 5637686
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2152 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2112
US-08-646-715-17

Query Match 1.1%; Score 48; DB 1; Length 2152;
Best Local Similarity 57.2%; Pred. No. 0.0031;
Matches 127; Conservative 0; Mismatches 90; Indels 5; Gaps 2;
QY 4034 GTCTTCAAGATGTTTCCAGCATCTGCAGGTTCTACATGGTCAGTATTATGACGGGA 4093
DB 1566 GCCTTTAAGATATTTCGCGGCATCTTGCTGATGGAATTGTACCAGATTCCATCCAA 1625
QY 4094 ATCAGAGATAGCAGTTGGAGCTCGCC--ATGGTTTCAAGTGGCCCTGTACGACATCGGAC 4151
DB 1626 TTCTAATTATGTTGCTACGGGCTCTGCAGACAGACTGTGGGCTCTGGAGCTCTGAA 1685
QY 4152 TGGAAATGT---CAGACATCCATGGACACAAAGGACCAATCACTGCACTGCTTTTC 4208
DB 1686 TGGTAATGTTGAAGGATCTTCACTGGACACAGGGACCAATTTCATTCCTTGACATTTTC 1745
QY 4209 TCTGTATGGAAGATATCTTGGCACCCTACTCAAAACACTGACAG 4250
DB 1746 TCCCAATGGAGATCTCTGGCTACAGGACCAACAGATGGCAG 1787

RESULT 5
US-09-806-708B-22
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT APPLICATION NUMBER: US/09/806,708B
CURRENT FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patent in version 3.0
SEQ ID NO 22


```
QY 2333 CTGAATATCGTCCAGCAAAATCAAGCCATTGACCCCTATTAGAAATATAATTTAACTATCG 2392
Db 1266 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1207

QY 2393 ACACCTCAAAAG 2403
Db 1206 RRRRRRRRRR 1196

RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAE1 promoters
US-09-806-708B-22

Query Match 0.9%; Score 39.2; DB 4; Length 1141;
Best Local Similarity 9.1%; Pred. No. 0.96;
Matches 72; Conservative 302; Mismatches 420; Indels 1; Gaps 1;

QY 930 TGTAACAATATCTCTTGGATCGAAAGATAAGAGTTGCTAATTGTCTCTCTTAC 989
Db 890 HSKRRTRHTRTCRTKYNNNNNNARTVYHHAARRWNNWTRTNNNNNNNNNACRN 831

QY 990 TCGGTTCTCTGATGACAGAAATATTTCAGTAACTCTTAATTGAGGTGATCTTC 1049
Db 830 TRTWABKHSWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 771

QY 1050 TGGAGGTTCAATATTTGGAACATATCAGACAGCTGATAAAGAGGAGTGAAGAAG 1109
Db 770 HAAVTTHTDWCYKTWNTWYDMMTTMBTTTTTRNMTTSTNNNNNNNNNNNNNN 711

QY 1110 GCTGGCAATGACAACTCTATTAGTTTGGAGAGGCAATTTGATAAAGTGAATCCTGTCC 1169
Db 710 MWKAYVAHANWGCWNNNTDABRTNNTTVMRRRWNTKTRWYSTTRHHYTGATNNNN 651

QY 1170 TGCTGGAATATAGATCAGCTGATGATTCCTCAATAGTAACTCTTAAAGTAAC 1229
Db 650 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 591

QY 1230 TGCAAGTGTGTACATACAGACATGACGACTTGTGTGCTGCTGAAGATGAAGCAT 1289
Db 590 NWMCRDVTYTRNNYCKSYAHSYVWYSSNNWYRRYSARNWSSMARWTTTRNNWMBV 531

QY 1290 AGTTATTGTACCTGCCACAGAGCGCCATAGTACAGCTGTTGCAAGGGGAAACACATGCT 1349
Db 530 RMRWAGTMMWRHNNNNNTDRTYVWVWKRARBTTTVYDSMCAKSMWRGNWWRAMWMA 471

QY 1350 CAGAAGGTTGGCCACTCAGAACACTCCGTTGCTGCGGACAAAGTCAATGTTT 1409
Db 470 ANNDAGANDHWTYMGNNTHWRRWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 411

QY 1410 GCTATATCTCATCAGCTCTCAGCTCGGTATGATCAAGATACTGATATCTGAGGTGT 1469
Db 410 KYMWKAAACNNNNBKMVRVAMWYSRDTTNTDMMWTSDBWHWYTDYTMVRAMNNNN 351
```

```
QY 1470 GGATTTTTCAGTCATAAATTTGGGACATATTTTCTGGAGAAATGAACATATCTTCTGTGT 1529
Db 350 NNNWBECKTTSWMMWHDHNTHTCTYGNNTWGSAYEWAAMSWAAGASNBVTYNWCWRMTY 291

QY 1530 TCATCGTGGTGAATTAATCAACTTCTAGTTCACCTGAAACTGTA-GTCAAGAGTAC 1588
Db 290 MGKWTNNNNNNKAWYRTKTVAWCNNRYDYDTAVMTBKRNKYCYAVBWWYBMYMGKH 231

QY 1589 AGCACTGCATCTGCTCTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1648
Db 230 WEWRRRABHRSWNNWVVKCRNKYMYSHYHAKWBAVGCNNNNWKKDRMAHHWCATN 171

QY 1649 AAAAATGCATAATGTTGGCATCTCTGACCTTTTCTTATTTCAAGTAAATCAATGAGGC 1708
Db 170 NNNMMWYAYMMHMKKGAATNNKTABRDDHBAHVKYTWYRVDYWCAMCMWNAKAV 111

QY 1709 CTTCGTGATGATPACC 1723
Db 110 RTAMKHMWYTTDRYV 96

RESULT 8
US-09-949-016-16919/c
; Sequence 16919, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16919
; LENGTH: 153866
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16919

Query Match 0.9%; Score 39.2; DB 4; Length 153866;
Best Local Similarity 62.0%; Pred. No. 32;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 4144 ATCCGAGCTGGAATAATGTCAGACATCCATGACACAGGACCAATCACTGAGTGGCT 4203
Db 20022 ATACATATCTGTTGTCTAGGAGAGTATTAAATGGAGAGGAACTGTTCAATAAAGTCGT 19963

QY 4204 TTGTCTCTGATGAAGATATCTTGCACCTACTCAAAACA 4243
Db 19962 GGTTCCTCAGAGGGGAAAATCTGACATCTACTCAAAACA 19923

RESULT 9
US-08-956-171B-52
; Sequence 52, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
```



```
;
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14

Query Match 0.9%; Score 38.2; DB 1; Length 7218;
Best Local Similarity 3.3%; Pred. No. 7.2;
Matches 10; Conservative 172; Mismatches 125; Indels 0; Gaps 0;

QY 2630 AGGAACTTACGAGTGTCCGCGTCACACAGACATCTCTGTCTATCAATTTCTT 2689
Db ||||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 AGGAGCTGCGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1112
QY 2690 TGGCAATATCTTAATGAGTATGACCAATGCAACTTTTATTGTGATCATATGAAGAAG 2749
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1113 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1172
QY 2750 GTCTTACCGCCAGCCTAGACCAAGCACCCACACCTTTCTAGGCAAGGGTTCCTC 2809
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1173 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1232
QY 2810 CAATCTCCAGTATATGTGCAAGGACAGATTAACAAGTGTGTCACCTGTCTTCG 2869
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1233 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1292
QY 2870 CTGGTCTGATGATGATCACTCTGGCTCTGACCTCTCTCTGCTCTCTTACATACCT 2929
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1293 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1352
QY 2930 GTTCTT 2936
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1353 YYYYYY 1359

RESULT 12
US-08-956-171B-230/c
; Sequence 230, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Pannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
```

```
;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 230:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2072 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 230:
US-08-956-171E-230

Query Match 0.8%; Score 38; DB 4; Length 2072;
Best Local Similarity 53.3%; Pred. No. 3.4;
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3322 CCACAAATGAAAAAATTTCTACATCTTACGAGGAAAGCAGCAAGCTACCGCTATT 3381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1086 CCAAAAATGACATATTAATGCAATAAAGCGCATATAATTAAGTAACCTTCTAACTGTT 1027
QY 3382 GTTTTACTTGGAGTAGTAGGAGCTGAATTTTGTGCTGAAATTTGAACCTCTCTAACTATTG 3441
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1026 GTTTCGCAACCAATTAATCCCGCTATCGTGTGTCTATAAATGAAGAACTCTGTACCG 967
QY 3442 ACCGACCTCGAAGCTCTAGCCAAATTCCT 3471
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
966 ATATTTCCTATACCATATAATACCATTTGCT 937

RESULT 13
US-08-781-986A-230/c
; Sequence 230, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
```


; INFORMATION FOR SEQ ID NO: 230:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2072 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-781-986A-230

Query Match

Best Local Similarity 0.8%; Score 38; DB 4; Length 2072;

Mismatches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3322 CCACAAATGAAGAAATTTTACATCTTACGAGGAAGCGGAGCAGCTACCGCTATT 3381

DB 1086 CCAAAATGACATAATAATGCAATGAAGCGAATAATAATTAAGTAACTTCTAAGCTTT 1027

QY 3382 GTTTTACTTGGAGTAATAGAGCTGCAATTTGGTGTCTGAAATGCACTCTAAACTATTG 3441

DB 1026 GTTTGCCAACCAATAATACCGCTATCGGTGTGCTAAATAATGAAGAACTGCTGTACCG 967

QY 3442 ACCAGACTCGAAGCTCTAGCCAAATTCCT 3471

DB 966 ATATTTCCCATACCAATAATACCATTTGCT 937

RESULT 14

US-10-071-411A-63

; Sequence 63, Application US/10071411A

; Patent No. 6797475

; GENERAL INFORMATION:

; APPLICANT: Joanne Barnes

; TITLE OF INVENTION: Detection of Polymorphisms in the Human

; FILE REFERENCE: MRI-021

; CURRENT APPLICATION NUMBER: US/10/071,411A

; CURRENT FILING DATE: 2002-02-07

; PRIOR APPLICATION NUMBER: 60/267,515

; PRIOR FILING DATE: 2001-02-08

; PRIOR APPLICATION NUMBER: 60/314,248

; PRIOR FILING DATE: 2001-08-21

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 63

; LENGTH: 168174

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(168174)

; OTHER INFORMATION: n = A,T,C or G

US-10-071-411A-63

Query Match

Best Local Similarity 0.8%; Score 38; DB 4; Length 168174;

Mismatches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 609 AATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAATTGCCAAGCATCTCTTTTG 668

DB 40595 ACTATTTTGTGATGTTGTAATTAATTAATTTCTTAATACCTTTGGATTGTTATAG 40654

QY 669 TGCATTTTACAAAGGTCACCTTTTGGTTGTGTGT 702

DB 40655 TTAATTTATAGAAAGGCACTGATTTTGTGTGT 40688

RESULT 15

US-10-071-411A-2

; Sequence 2, Application US/10071411A

; Patent No. 6797475

; GENERAL INFORMATION:

; APPLICANT: Joanne Barnes

; APPLICANT: Joanne Meyer

; TITLE OF INVENTION: Detection of Polymorphisms in the Human

; FILE REFERENCE: MRI-021

; CURRENT APPLICATION NUMBER: US/10/071,411A

; CURRENT FILING DATE: 2002-02-07

; PRIOR APPLICATION NUMBER: 60/267,515

; PRIOR FILING DATE: 2001-02-08

; PRIOR APPLICATION NUMBER: 60/314,248

; PRIOR FILING DATE: 2001-08-21

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 168273

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(168273)

; OTHER INFORMATION: n = A,T,C or G

US-10-071-411A-2

Query Match

Best Local Similarity 0.8%; Score 38; DB 4; Length 168273;

Mismatches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 609 AATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAATTGCCAAGCATCTCTTTTG 668

DB 40645 ACTATTTTGTGATGTTGTAATTAATTAATTTCTTAATACCTTTGGATTGTTATAG 40704

QY 669 TGCATTTTACAAAGGTCACCTTTTGGTTGTGTGT 702

DB 40705 TTAATTTATAGAAAGGCACTGATTTTGTGTGT 40738

Search completed: July 2, 2005, 13:01:55

Job time : 726 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 03:12:11 ; Search time 12549 Seconds
(without alignments)

13567.733 Million cell updates/sec

Title: US-10-645-335-1

Perfect score: 4473

Sequence: 1 atggcaggaaacagcctgt.....agcacgccttcgtgctaa 4473

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1843.2	41.2	2422	3	BC026618 Mus muscu
2	1279	28.6	2067	3	AK052730 Mus muscu
3	803.8	18.0	831	1	AU132411 AU132411
4	770.8	17.2	882	5	BUI188187
5	734.6	16.4	758	6	CD652969
6	670.6	15.0	678	4	B1549586
7	669.8	15.0	688	4	B1756477
8	648.4	14.5	650	5	BX507682
9	642	14.4	653	7	CN314767
10	619.6	13.9	766	4	B1855881
11	613.4	13.7	768	6	CA328908
12	608.2	13.6	780	5	BQ770731
13	601.4	13.4	603	5	BX955125
14	601.2	13.4	631	7	CR544624
15	593.2	13.3	714	7	CF535467
16	586	13.1	738	4	BF984064
17	577	12.9	613	1	AL042212
18	554.6	12.4	801	7	CK481829
19	553.8	12.4	693	6	BY763936
20	546.6	12.2	878	5	BX688519
21	544.2	12.2	625	2	BE007321
22	522	11.7	660	2	BF148762
23	513	11.5	652	7	CK636646
24	511.8	11.4	632	7	CF537838

RESULT 1
BC026618
LOCUS BC026618 2422 bp mRNA linear HTC 20-SEP-2002
DEFINITION Mus musculus, clone IMAGE:4977270, mRNA.
ACCESSION BC026618
VERSION BC026618.1 GI:20071833
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
REFERENCE 1 (bases 1 to 2422)
AUTHORS Direct Submission
TITLE Submitted (02-APR-2002) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

ALIGNMENTS

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 59 Row: k Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenesScan gene prediction
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..2422
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4977270"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."

/clone.lib="NCI_CGAP_Mam6"
/lab_host="DH105"
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 41.2%; Score 1843.2; DB 3; Length 2422;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 263; Indels 0; Gaps 0;
QY 1 ATGGCAGGAAACAGCCTTGTCTACCCATTGTTCTTTGGGTGAAAGCGGCCACACAT 60
DB 158 ATGGCAGGAAACAGCCTTGTCTGCCCATTTGTTCTTTGGGCGCGCAAGCACCACACAT 217
QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCTTAACAGAGTGTCCAGAC 120
DB 218 TGCATTTCTGTCATACTGTTGACAGATGATGGGGGACAAATTTAACTGATGCCACGAT 277
QY 121 GGACAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAATTAATCTCCAGACATGTTG 180
DB 278 GGACAAATATGTCTCTGGGATGTTTCGGTAGAACTAGNAGTTAATCCCGAGCACTGTTA 337
QY 181 TTTGGTCATACAGCATCAATCACTTGTGTTCTTAAAGCTTGTGCTTCAGTGACAAACAG 240
DB 338 TTTGGCCACACAGCATCCATCACTTGTGTTGTCAAAGCCCTGCGCTTCTGGGACAAAGCG 397
QY 241 TATATTGTGATGCATCTGAAAGTGGAGATGTCCTCTGGATGTGATGTGATGGCAGA 300
DB 398 TACACTGTGAGCGGTCTGCAAAACGAGAGATGTGCTCTGGATGTGAAACGATGGGAGA 457
QY 301 TGTATTGAAATTTACAAATTTAGCTTGCACACATCTGGCATACAGTTCTACCAAGTTCTCT 360
DB 458 TGTATTGAAATTTACCAAGTTAGCTTGCACACATCTGGCATACAGTTCTACCAAGTTCTCT 517
QY 361 GTTGGGAATCAGCGAAGAAAGGCTTTTATGCGACGACATTAACCTGAAATTCCTTGT 420
DB 518 GTTGGGAACAGCAAGAGGCGAGCTCTCTGCCATGGACATTAACCTGAAATTCCTCGTT 577
QY 421 GTGGATGCTACAGCCTTGAAGTATTATCTCTTAGTATCAAGATATACCCAGACTGG 480
DB 578 GTGGATGCCACAGCCTTGAAGTATTATCTCTTAGTATCAAGATATCTCTCAGACTGG 637
QY 481 ATTAGTCTCATAGTATTTATTCGATCCCAACGACACAGAGGACACAGTGTAGCACTC 540
DB 638 ATCAGCTCCATGAGCATCATCCGCTCTCAGCGGACACAGAGGACACTGTGTGGCGCTC 697
QY 541 TCGGTGATCGGCATCTCTGAAGGCTGGAATTTGTTTACCTCGGAAATTAAGTGACATGCGAG 600
DB 698 TCTGTGACAGGTATCTGAAGGTGTGGATTTGTCACTCTGAAATGAGTGGAAATGCAAGAT 757
QY 601 ACTGAGCCATATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCGCAAGCATC 660
DB 758 ACTGAGCCATATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCGCAAGCATC 817
QY 661 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTTCCAAATATTCGAGGGTG 720
DB 818 TCTTTTGTGCAATTTACACAGAGGTCACTTTTGGTGTGTGCTCCAAATACTGAGGGTG 877
QY 721 TTCGATGCCGGAGCATTTCTCTTGTGTGTTCAGGTCCTAGTGAAATGAGACAGACATGG 780
DB 878 TTTGATGCTGGGCACTACTCTGTGTGTGCTCAGGTCCTAGTGAATAATGACACAGACATGG 937
QY 781 ACCGGGGGACATTTGTCATCAGATAAGTGCATCATTTGGACAGAAATGGCAAGT 840
DB 938 ACTGGAGGGGACATTTGTGTCTGACAGAAAGTCAATCATCTGGACAGAAATGGCAGAGT 997
QY 841 TATATTTACAAACTACTGCGCAGTTGCCITTCAGCTAGTGTATTCATTCGCGAGTGTGTG 900
DB 998 TACATCTACAACTCCCTGCCAGTTGCCITTCAGCTAGTGTATTCATTCGCGACGAGCTG 1057
QY 901 GGGAAAGCAGTTGAAATTTTAAATTCCTCTGTGTAACAATATCCCTTTGGATCGAAAGAT 960
DB 1058 GGGAAAGCAGTTGAAATTTTAAATTCCTCTGTGTCAGCATAGCCTCTTTGGATCAGAAAGAT 1117

QY 961 AAAGAGTTCTCTAATTTGTTCTCTCTGTTACTCGGTTCTTCTATGATGACAGAAATATTTTC 1020
DB 1118 AAAGAGTTGGTAATTTGTTCTCTCTGTTACTCGGTTTCTACGGAATGCAAGGAATATTG 1177
QY 1021 CATAAACTGTTAATTTACGGGTGATTTCTTCTGGAAGTTGAAATATTTTGGAAACATATCAGAC 1080
DB 1178 CATAGCTACTAATTTACGGGCGATTTCTTCTGGAAGTTAAATATCTTGGAAACATATGACAGAC 1237
QY 1081 ACAGCTGATAAACAGGGAAGTGAAGAGGCTGGCAATGCACAACTTCTATTAGTTTGGCAA 1140
DB 1238 ATAGCAGAGAAACAGGAAGCCGATGAAGGGCTAAAGATGACAACTTTGATATTAGCTTGGCAA 1297
QY 1141 GAGGCAATTTGATAAACTGAATCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGAT 1200
DB 1298 GAGGCAATTTGACACAGCTGAAGCCTTGGCCTCTGGBAATTTATCGATCAGCTAAGTGTGAT 1357
QY 1201 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTATACACGACATGACGA 1260
DB 1358 CCCAACAGCAACGAACTCTTAAAGTAACTGCGAGTGTCTACATACAGCACACGCGCGA 1417
QY 1261 CTGTTTGTGCTCGTGAAGATGGAGCATAGTTATTTGTACCTGCGCACACAGACGGGCATA 1320
DB 1418 CTGTTTGTGCGCGGGAAGATGGAAGCATCATTTTGTGCTGCGCACCCAGACGGGCATA 1477
QY 1321 GTACAGCTGTGCAAGGGGAAACACATGCTCAGAGAGGTTGGCCACTCACAGAAACATCTC 1380
DB 1478 GTTACGTAATTTGAGGAGAAACACATGCTCAGAGAGGTTGGCCCTCACAGAAACCTC 1537
QY 1381 CGTGTCTATCGGAAACAAAGTCACTGTTTGTCTATATCTCATCAGGTCTCAGCTCGGTAT 1440
DB 1538 CGTGGCCATCGGAAACAAAGTCACTGCTGCTGTATCTCCCATCAGGTCTCAGCTCGGTAT 1597
QY 1441 GATCAAGATACTGATATCTGAGGTGTGGAATTTTTCAGTCAATATTTGGGACATATTT 1500
DB 1598 GACCAAGATATCTGATATCTGAGGTGTGGAATTTTTCGTCATATTTTGGGACATATTT 1657
QY 1501 TCTGGAATAATGAAACATATCTTCTGTGTTTCATGTTGTGTGAGTACTCAACTCTTAGTT 1560
DB 1658 TCTGGGAATTTGAACATATCTTCTGTGTTTCATGTTGTGTGAGATCACACAACTCTTGGT 1717
QY 1561 CCACTCTGAAATCTGTATGCAAGATPACAGCATGTCATCTGCTCTGTAGCGATGACAC 1620
DB 1718 CCCCCGAAATCTGTAGTCAAGAGTTCAACATGTCATCTGTTCCGTAGCAGTACGAC 1777
QY 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCATATGTTGGCATCTGTCACCTT 1680
DB 1778 TCTGTAGGCTCTCTAAGCTCTGGAGAGAAATGATCATGTTGGCGTCTGCCACCTC 1837
QY 1681 TTTCTATTCAAGTAATCAATGGAGGCTTCTGATGATTAACCTGTGGTGGGCTGTCTCA 1740
DB 1838 TTTCTATTTCAGGTGATCAAGTGGAGGCTTCCGACGACTACTCTGTGTGGATGACACA 1897
QY 1741 GATGTTCTGTGTACGTTCTGGCAATPAGATACGTGTGCAATGGATCGTTGTGTGTGAGGG 1800
DB 1898 GACGGCTCTGTGTATGCTGTGGCAGATGACACTGTGTGCGCTGGATCGTTGTGCAATGGGA 1957
QY 1801 ATAAACAGCATTTGAGATTTCTAAACGCTTCTGTATGAGCTGTTCCCTGCTGTGTGATTTCA 1860
DB 1958 ATCACAGCAGTGGAGATCTGAACGCTTGTGAAGAGCGGTTCTCCCGCTGTGAGACTCT 2017
QY 1861 CTTAGTCTACCGACAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGTGCTCTT 1920
DB 2018 CTTAGTCTACCGACAGTCAACCTTGAACCAAGCCATGACAGAGGAGTCTCGCGGCCCTT 2077
QY 1921 AAAAATATCGGCCCATCATTAAGCTA CAAACCTTTGCAACTAAACCTCTTGGCTTCTGAGGCA 1980
DB 2078 AAAAATATCGGCCCATCCATTAAGCTGCAAAACCTTTGCAACGAAACCTTTTGGCTTCTGAGGCC 2137
QY 1981 TCTGCAAGAGGGAATTTTACATAATTTCTCAATCTCCCTGATGTTTCAAGCAATAAAG 2040
DB 2138 TCTGCAAGAGGGAATTTTACATAATTTCCATACTCCCTGATGTTTCAAGCAATAAAG 2197
QY 2041 ACAAACCTTAACAGACCCGACACATACATGTGTCTATTTTGTGATGTGGAAGCGTTGATTTT 2100

Db 1 CTATTCAAGTAATCAATGGAGGCTTCTGATGATTACCTGGTGGGTGTTTCAGATG 60
Qy 1745 GTTCTGTGTACGTCTGGCAATGATATCTGTGTGATTCGATTCGTTGTGATGGGATAA 1804
Db 61 GTTCTGTGTACGTCTGGCAATGATATCTGTGTGATTCGATTCGTTGTGATGGGATAA 120
Qy 1805 CAGCAGTTGAGATTTAAACGCTGTGATGAGCTGTTCCTGCTGCTGTTGATTCACCTTA 1864
Db 121 CAGCAGTTGAGATTTAAACGCTGTGATGAGCTGTTCCTGCTGCTGTTGATTCACCTTA 180
Qy 1865 GTCATCCAGCAGTCAAACTAAACCAAGCTATGACGACGCTAGTCTTGTGCTCTTAAAA 1924
Db 181 GTCATCCAGCAGTCAAACTAAACCAAGCTATGACGACGCTAGTCTTGTGCTCTTAAAA 240
Qy 1925 ATATGGCCCATATAAGCTCAAAACCTTGGCACTTAACCTCTTGGCTTCTGAGCAGTCG 1984
Db 241 ATATGGCCCATATAAGCTCAAAACCTTGGCACTTAACCTCTTGGCTTCTGAGCAGTCG 300
Qy 1985 ACAAGGGAATTTACCTAAATATTTCTCAATCTCCCTGATGGTTCAGGCAATTAAGCAA 2044
Db 301 ACAAGGGAATTTACCTAAATATTTCTCAATCTCCCTGATGGTTCAGGCAATTAAGCAA 360
Qy 2045 ACCTAACAGACCCGACATACATGCTGCTATTTCTTGATGGAAGCTTGCATTATTCAC 2104
Db 361 ACCTAACAGACCCGACATACATGCTGCTATTTCTTGATGGAAGCTTGCATTATTCAC 420
Qy 2105 TCCTGACTGAGAGCCCTTAGGCGGAATACTGCTCTTATTTCCCAAGAGATTTGCAA 2164
Db 421 TCCTGACTGAGAGCCCTTAGGCGGAATACTGCTCTTATTTCCCAAGAGATTTGCAA 480
Qy 2165 AAGCATCTGCAGTTTACAGAAAGGGGCTCTTTTAACTGGAAGAGAGAGAGAGTTC 2224
Db 481 AAGCATCTGCAGTTTACAGAAAGGGGCTCTTTTAACTGGAAGAGAGAGAGAGTTC 540
Qy 2225 TCTTCCAAAGTGAAGAACCATCAAGAGACATCAAGAGACACCTCTCTGATGATG 2284
Db 541 TCTTCCAAAGTGAAGAACCATCAAGAGACATCAAGAGACACCTCTCTGATGATG 600
Qy 2285 AAGAGGAGGATGAGGAGATTAATGAGGAGAGAGAGAGAGAGAGTCTCTGATGATG 2344
Db 601 AAGAGGAGGATGAGGAGATTAATGAGGAGAGAGAGAGAGAGAGTCTCTGATGATG 660
Qy 2345 CAGGCAATCAAGGCAATGAGCCTTATGAGATTAATTAATTAATGAGACATGCAAGAGC 2404
Db 661 CAGGCAATCAAGGCAATGAGCCTTATGAGATTAATTAATTAATGAGACATGCAAGAGC 720
Qy 2405 TGTATTATGCTCCTGCTTACGCTTGGGTTTGAATGAGTACTGATGAGTGTGCTGG 2464
Db 721 TGTATTATGCTCCTGCTTACGCTTGGGTTTGAATGAGTACTGATGAGTGTGCTGG 779
Qy 2465 ATGCTCTTGAATGCTGAAACCCCACTGACCCCTATGCTTGTGCTCTTG 2514
Db 780 ATGCTCTTGAATGCTGAAACCCCACTGACCCCTATGCTTGTGCTCTTG 829

RESULT 4

BU188187 882 bp mRNA linear EST 04-SEP-2002
LOCUS AGENCOURT_7928616 NIH_MGC_72 Homo sapiens cdna clone IMAGE:6151347
DEFINITION 5', mRNA sequence.
ACCESSION BU188187
VERSION BU188187.1 GI:22702171
KEYWORDS EST.
SOURCE Homo sapiens. (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLMI3487 row: h column: 04
High quality sequence stop: 477.
Location/Qualifiers
source 1..882
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6151347"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_72"
/notes="Organ: Skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2 kb. Library constructed by Life Technologies." ORIGIN
Query Match 17.2%; Score 770.8; DB 5; Length 882;
Best Local Similarity 98.8%; Pred. No. 3.1e-211;
Matches 808; Conservative 0; Mismatches 7; Indels 3; Gaps 3;
Qy 3573 TCCTCCACACAGCAGCTATCCGAGACAGCCATTCATGATTCGATGGGGTTCCTGCAATCTTGTGCGGA 3632
Db 3 TTCTCTCCACAGCAGCTATCCGAGACAGCCATTCATGATTCGATGGGGTTCCTGCAATCTTGTGCGGA 62
Qy 3633 TTGGGAGCCTTACATGATGATGTCGCTGTCGATGGGGTTCCTGCAATCTTGTGCGGA 3692
Db 63 TTGGGAGCCTTACATGATGATGTCGCTGTCGATGGGGTTCCTGCAATCTTGTGCGGA 122
Qy 3693 TGCCGAGAAACAACCTTGCACACATCAAAATGGGGTTCCTGAGCCAGCAGCTGATC 3752
Db 123 TGCCGAGAAACAACCTTGCACACATCAAAATGGGGTTCCTGAGCCAGCAGCTGATC 182
Qy 3753 GGCCCGCTCTGGAGGAGTACACAGACATACGCTCTTCGATGGGGTTCCTGAGCCAGCAGCTGATC 3812
Db 183 GGCCCGCTCTGGAGGAGTACACAGACATACGCTCTTCGATGGGGTTCCTGAGCCAGCAGCTGATC 242
Qy 3813 CACACATAGCCAAAGAGGTACACAGACATACGCTCTTCGAGCAATAACCAATCACA 3872
Db 243 CACACATAGCCAAAGAGGTACACAGACATACGCTCTTCGAGCAATAACCAATCACA 302
Qy 3873 GCAGAAATATGCACACAACTCTTCGAGGCTAAAGGGGAAATTTTGAGAGTCAATGA 3932
Db 303 GCAGAAATATGCACACAACTCTTCGAGGCTAAAGGGGAAATTTTGAGAGTCAATGA 362
Qy 3933 AATTCCTATGAAAGATGCCACAGATGTTGGATCTTCGATGGAGTTATGACAT 3992
Db 363 AATTCCTATGAAAGATGCCACAGATGTTGGATCTTCGATGGAGTTATGACAT 422
Qy 3993 CATATGATCTGCTTGAAGGATCTTTAGTTAAAAAGAAAGGTCTTCAAGAAATGTTTCCC 4052
Db 423 CATATGATCTGCTTGAAGGATCTTTAGTTAAAAAGAAAGGTCTTCAAGAAATGTTTCCC 482
Qy 4053 AGCCATCTGAGGTTTACATGCTCAGCTATTTATGAGCGGAATCAGAGATAGCAGTTGG 4112
Db 483 AGCCATCTGAGGTTTACATGCTCAGCTATTTATGAGCGGAATCAGAGATAGCAGTTGG 542
Qy 4113 AGCTGCCATGTTTACGTCGCCCTTACACATCCGAGCTCGAAATGTCAGCAATCCA 4172
Db 543 AGCTGCCATGTTTACGTCGCCCTTACACATCCGAGCTCGAAATGTCAGCAATCCA 602
Qy 4173 TGGACACAGGAGCAATCACTGAGTGGCTTTGCTCTGATGGAGATATCTTCCAC 4232
Db 603 TGGACACAGGAGCAATCACTGAGTGGCTTTGCTCTGATGGAGATATCTTCCAC 662
Qy 4233 CTACTCAACACATGACAGCCACATTTCTTTTGGCAGATGAAACGCTCACTGCT-GGGAA 4291

Tissue Procurement: ATCC/DCTD/DP

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLMI3487 row: h column: 04

High quality sequence stop: 477.

Location/Qualifiers

FEATURES

source

1..882

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6151347"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC_72"

/notes="Organ: Skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 17.2%; Score 770.8; DB 5; Length 882;

Best Local Similarity 98.8%; Pred. No. 3.1e-211;

Matches 808; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Qy 3573 TCCTCCACACAGCAGCTATCCGAGACAGCCATTCATGATTCGATGGGGTTCCTGCAATCTTGTGCGGA 3632

Db 3 TTCTCTCCACAGCAGCTATCCGAGACAGCCATTCATGATTCGATGGGGTTCCTGCAATCTTGTGCGGA 62

Qy 3633 TTGGGAGCCTTACATGATGATGTCGCTGTCGATGGGGTTCCTGCAATCTTGTGCGGA 3692

Db 63 TTGGGAGCCTTACATGATGATGTCGCTGTCGATGGGGTTCCTGCAATCTTGTGCGGA 122

Qy 3693 TGCCGAGAAACAACCTTGCACACATCAAAATGGGGTTCCTGAGCCAGCAGCTGATC 3752

Db 123 TGCCGAGAAACAACCTTGCACACATCAAAATGGGGTTCCTGAGCCAGCAGCTGATC 182

Qy 3753 GGCCCGCTCTGGAGGAGTACACAGACATACGCTCTTCGATGGGGTTCCTGAGCCAGCAGCTGATC 3812

Db 183 GGCCCGCTCTGGAGGAGTACACAGACATACGCTCTTCGATGGGGTTCCTGAGCCAGCAGCTGATC 242

Qy 3813 CACACATAGCCAAAGAGGTACACAGACATACGCTCTTCGAGCAATAACCAATCACA 3872

Db 243 CACACATAGCCAAAGAGGTACACAGACATACGCTCTTCGAGCAATAACCAATCACA 302

Qy 3873 GCAGAAATATGCACACAACTCTTCGAGGCTAAAGGGGAAATTTTGAGAGTCAATGA 3932

Db 303 GCAGAAATATGCACACAACTCTTCGAGGCTAAAGGGGAAATTTTGAGAGTCAATGA 362

Qy 3933 AATTCCTATGAAAGATGCCACAGATGTTGGATCTTCGATGGAGTTATGACAT 3992

Db 363 AATTCCTATGAAAGATGCCACAGATGTTGGATCTTCGATGGAGTTATGACAT 422

Qy 3993 CATATGATCTGCTTGAAGGATCTTTAGTTAAAAAGAAAGGTCTTCAAGAAATGTTTCCC 4052

Db 423 CATATGATCTGCTTGAAGGATCTTTAGTTAAAAAGAAAGGTCTTCAAGAAATGTTTCCC 482

Qy 4053 AGCCATCTGAGGTTTACATGCTCAGCTATTTATGAGCGGAATCAGAGATAGCAGTTGG 4112

Db 483 AGCCATCTGAGGTTTACATGCTCAGCTATTTATGAGCGGAATCAGAGATAGCAGTTGG 542

Qy 4113 AGCTGCCATGTTTACGTCGCCCTTACACATCCGAGCTCGAAATGTCAGCAATCCA 4172

Db 543 AGCTGCCATGTTTACGTCGCCCTTACACATCCGAGCTCGAAATGTCAGCAATCCA 602

Qy 4173 TGGACACAGGAGCAATCACTGAGTGGCTTTGCTCTGATGGAGATATCTTCCAC 4232

Db 603 TGGACACAGGAGCAATCACTGAGTGGCTTTGCTCTGATGGAGATATCTTCCAC 662

Qy 4233 CTACTCAACACATGACAGCCACATTTCTTTTGGCAGATGAAACGCTCACTGCT-GGGAA 4291

Db 663 CTRACTAAACACTGACGACACATTTCTTTTGGCAGATGAACACGCTCACTGCTGGGAA 722
QY 4292 GCATCGGCGATGTAACCTGGCACTCAGCTGCGCTGCATTAATAAACCCTACCGAGTGCCTCC 4351
Db 723 GCATCGGCGATGTAACCTGGCACTCAGCTGCGCTGCATTAATAAACCCTACCGAGTGCCTCC 782
QY 4352 CT-GTGCAGCCCGCTGCTCCCTCCGCTCCCAATGATGCC 4387
Db 783 CTGGTGCAGCCCGCTGCTCCCTCCGCTCCCAATGATGCC 820

RESULT 5
CD652969 758 bp mRNA linear EST 18-JUN-2003
LOCUS AGENCOURT_14541645 NIA Human H1 Embryonic Stem Cell cDNA Library
DEFINITION (Long) Homo sapiens cDNA clone IMAGE:30420765 5', mRNA sequence.
ACCESSION CD652969
VERSION CD652969.1 GI:31890810
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 758)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Daniela S. Gerhard, Ph.D.
COMMENT Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
cDNA Library Preparation: Yulan Piao and Minoru KO
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC cDNA library
Can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM497 row: h column: 22
High quality sequence stop: 719.
Location/Qualifiers
1. 758
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30420765"
/tissue_type="Embryonic Stem cells"
/cell_line="WA01"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library (Long)"
/notes="Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI;
This is a long-transcript enriched cDNA library (Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199] from WA01 cell line. Undifferentiated human ES cell line WA01/H1 was obtained from WiCell Research Institute, Inc., Madison, WI, cultured according to their instructions, on ME9 feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTP, TERT, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCAM, MSX1, Flt3, SSEA-1, TUBB3, NES, GFAP, and BOMES. When confluent (18-10 days after plating), the ES cells from 4 x 6 cm dishes were treated with 1 mg/ml collagenase, type IV
(Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with TRIzol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID:11544199] Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTCTAGATCGCGCGCGCTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and

FEATURES
source

ORIGIN

Query Match 16.4%; Score 734.6; DB 6; Length 758;
Best Local Similarity 99.5%; Pred. No. 9.3e-201;
Matches 737; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2939 TAAATGAAGTGTGGAGTCAGTTAGCTGTATGCTGTATGCTGCGAGACTACTGG 2998
Db 17 TTAACAAGTGTGGAGTCAGTTAGCTGTATGCTGTATGCTGCGAGACTACTGG 76
QY 2999 GATTGATTAATTTAGGCTCTCCCTTCTGAGATGCTGCGCGAGATGCGAGATCGAT 3058
Db 77 GATTGATTAATTTAGGCTCTCCCTTCTGAGATGCTGCGCGAGATGCGAGATCGAT 136
QY 3059 GCTTGGAGTGTAGAGAAAGCGCACAGGCCCTGCTTCTGCGGAACTGAGAAGAAATTGAGC 3118
Db 137 GCTTGGAGTGTAGAGAAAGCGCGAGGCCCTGCTTCTGCGGAACTGAGAAGAAATTGAGC 196
QY 3119 AGCGAGCGAGGAAGCAATTTGATGCTGCGGCTCTTCTTACCTTACAGTACATAGACC 3178
Db 197 AGCGAGCGAGGAAGCAATTTGATGCTGCGGCTCTTCTTACCTTACAGTACATAGACC 256
QY 3179 AGCTCATATCCTGAGTCAATCAGAAAGCGCGAGACTATCACCAGGCTCTCTGATG 3238
Db 257 AGCTCATATCCTGAGTCAATCAGAAAGCGCGAGACTATCACCAGGCTCTCTGATG 316
QY 3239 CCTCAGGCGCTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTTGTGTGACGATGACATCA 3298
Db 317 CCTCAGGCGCTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTTGTGTGACGATGACATCA 376
QY 3299 CCACCTGCTTCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAA 3358
Db 377 CCACCTGCTTCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAA 436
QY 3359 GACGGAAGCAAGCTACCGCTATTTGTTTCTGGAGTAATGAGCTGAATTTGGTGTCTG 3418
Db 437 GACGGAAGCAAGCTACCGCTATTTGTTTCTGGAGTAATGAGCTGAATTTGGTGTCTG 496
QY 3419 AAATTCGAACCTCTTAAACTATTGACCAGACCTCGAAGCTCTAGCCAAATTTCTTGGAGGAT 3478
Db 497 AAATTCGAACCTCTTAAACTATTGACCAGACCTCGAAGCTCTAGCCAAATTTCTTGGAGGAT 556
QY 3479 TCGGGTTGACTAGTGTGGATCCAACTACTCTGCTGCGCAGACATATTTGCAAGGCACTGA 3538
Db 557 TCGGGTTGACTAGTGTGGATCCAACTACTCTGCTGCGCAGACATATTTGCAAGGCACTGA 616
QY 3539 CGTTTCTTCTGTACAGCTTCCAGGCCCAAACTTCTTCCACACAGACATATTCGAGAA 3598
Db 617 CGTTTCTTCTGTACAGCTTCCAGGCCCAAACTTCTTCCACACAGACATATTCGAGAA 676
QY 3599 CAGCCATTCATCTGATTGGACGTGGGTTCACCTGTTTGGGAGCCCTTACATGGATGTGTCCG 3658
Db 677 CAGCCATTCATCTGATTGGACGTGGGTTCACCTGTTTGGGAGCCCTTACATGGATGTGTCCG 736
QY 3659 CTGTTCTGATGGGCTTCTCG 3679
Db 737 CTGTTCTGATGGGCTTCTCG 757

RESULT 6
BI549586
LOCUS

BI549586 678 bp mRNA linear EST 05-SEP-2001

DEFINITION 603194003F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5265024 5',
mRNA sequence.
ACCESSION B1549586
VERSION B1549586.1 GI:15436898
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 678)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>.
Plate: LLAM1668 row: f column: 01
High quality sequence stop: 678.
Location/Qualifiers
1. .678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5265024"
/issue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC 95"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI
(gtcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.5 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this
is a NIH_MGC Library."

ORIGIN
Query Match 15.0%; Score 670.6; DB 4; Length 678;
Best Local Similarity 99.4%; Pred. No. 3.3e-182;
Matches 673; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2625 AGGAAGGGAACTTACGGAGTGTCCCGTCCGTCACACAGCATCTCCTGTCTATCAT 2684
DB 1 AGCGAGGGAACTTACGGAGTGTCCCGTCCGTCACACAGCATCTCCTGTCTATCAT 60
QY 2685 TTCTTTGGCAATACCTTTAATGAGTATGACCAATGCACTTTATTGGTGATCATATGAA 2744
DB 61 TTCTTTGGCAATACCTTTAATGAGTATGACCAATGCACTTTATTGGTGATCATATGAA 120
QY 2745 GAAGGTCTTACAGGGCCTAGACCAACACCCAGCATTTCTAAGCAAGGGGTTTC 2804
DB 121 GAAGGTCTTACAGGGCCTAGACCAACACCCAGCATTTCTAAGCAAGGGGTTTC 180
QY 2805 CCTTCCAACTTCCAGTAATATTGCAAGCAGATTAACAAGTGTCTGCACCTGTGCT 2864
DB 181 CCTTCCAACTTCCAGTAATATTGCAAGCAGATTAACAAGTGTCTGCACCTGTGCT 240
QY 2865 TTCGGTCTGATGCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTCTCTTTACA 2924
DB 241 TTCGGTCTGATGCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTCTCTTTACA 300
QY 2925 TACTGTGTTCTTAGTAATGAAGTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 2984
DB 301 TACTGTGTTCTTAGTAATGAAGTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 360

QY 2985 GCCAGACCTACTCGGATTTGGATTAATTTAGGCTCCCTCTCTGAGATGCTGCCCGAAG 3044
DB 361 GCCAGACCTACTCGGATTTGGATTAATTTAGGCTCCCTCTCTGAGATGCTGCCCGAAG 420
QY 3045 ATGGCAAGATCGATGCTTTGGAGTGAGAGAAGCGCACAGGCCCTGCTTCTGCGGAAC 3104
DB 421 ATGGCAAGATCGATGCTTTGGAGTGAGAGAAGCGCGAGGCCCTGCTTCTGCGGAAC 480
QY 3105 GAGAAGAATTTGACGAGCGAGGAGGAAGCAATTTGATGCTGGGCTCTTTACTTACC 3164
DB 481 GAGAAGAATTTGACGAGCGAGGAGGAAGCAATTTGATGCTGGGCTCTTTACTTACC 540
QY 3165 TCAGTACATAGACCACTGTCATATCATCTGAGTGACATCATGAAGCGCGAGACTATCAC 3224
DB 541 TCAGTACATAGACCACTGTCATATCATCTGAGTGACATCATGAAGCGCGAGACTATCAC 600
QY 3225 CACGGCTCTGATGCTCTAGGCGCTCAAGCACAAGTCCAGGAGGAGAGCATGACCTTGT 3284
DB 601 CACGGCTCTGATGCTCTAGGCGCTCAAGCACAAGTCCAGGAGGAGAGCATGACCTTGT 660
QY 3285 TGACGATGACATCACCA 3301
DB 661 TGACGATGACATCACCA 677
RESULT 7
BI756477
LOCUS 688 bp mRNA linear EST 25-SEP-2001
DEFINITION 603024602F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5195065 5',
mRNA sequence.
ACCESSION BI756477
VERSION BI756477
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 688)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1488 row: c column: 02
High quality sequence start: 3
High quality sequence stop: 688.
Location/Qualifiers
1. .688
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5195065"
/lab_host="DH10B"
/clone_lib="NIH_MGC 114"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."

FEATURES
source

ORIGIN

```
Query Match      15.0%; Score 669.8; DB 4; Length 688;
Best Local Similarity 99.6%; Pred. No. 5.6e-182;
Matches 682; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 3668 TGGGGCTTCGAACTTTGTGCGGATGCGCGAGAAACAATTCGCAACATCACAATGGGGT 3727
Db 1 TGGGGCTTCGAACTTTGTGCGGATGCGCGAGAAACAATTCGCAACATCACAATGGGGT 60

QY 3728 TGCCCTCTGAGCCAGCAGCTGACTCGGCCCGCTCTGCGAGGAGTCCCTCTGCTCATG 3787
Db 61 TGCCCTCTGAGCCAGCAGCTGACTCGGCCCGCTCTGCGAGGAGTCCCTCTGCTCATG 120

QY 3788 CCACCCGCCAGACCAACCGGCTTCATCACCACCATAGCCAAAGAGGTACACAGATACGG 3847
Db 121 CCACCCGCCAGACCAACCGGCTTCATCACCACCATAGCCAAAGAGGTACACAGATACGG 180

QY 3848 CTCTTCGAGCAATACCCAAATCAACAGAGAAATGACACAACTCTTGCACGAGCTA 3907
Db 181 CTCTTCGAGCAATACCCAAATCAACAGAGAAATGACACAACTCTTGCACGAGCTA 240

QY 3908 AAGGGGAATTTGAGAGTCATGAAATTTCTATTGAAAGATGCCACAGATGTTGG 3967
Db 241 AAGGGGAATTTGAGAGTCATGAAATTTCTATTGAAAGATGCCACAGATGTTGG 300

QY 3968 ATCTTCTGAGAGTTATGGACATCATATTATGCTACTGCGCTTGAAGGATCTTTAGTTAAA 4027
Db 301 ATCTTCTGAGAGTTATGGACATCATATTATGCTACTGCGCTTGAAGGATCTTTAGTTAAA 360

QY 4028 AGAAGGCTTCAAGAAATGTTTCCAGCCATCTGCAAGGTTCTACATGTCAGCTATTATG 4087
Db 361 AGAAGGCTTCAAGAAATGTTTCCAGCCATCTGCAAGGTTCTACATGTCAGCTATTATG 420

QY 4088 AGCGGAATCACAGAAATAGCAGTTGGAGCTCGCCATGTTGAGTGGCCCTGTACGACATCC 4147
Db 421 AGCGGAATCACAGAAATAGCAGTTGGAGCTCGCCATGTTGAGTGGCCCTGTACATACCC 480

QY 4148 GGACTGGAATGTCAGACAAATCCATGGACACAAAGGACCAATCACTGCAAGTGGCTTTTG 4207
Db 481 GGACTGGAATGTCAGACAAATCCATGGACACAAAGGACCAATCACTGCAAGTGGCTTTTG 540

QY 4208 CTCTGATGGAAGATATCTTGCCACCTACTTCAAACTGACAGCCACATTTCTTTTGGC 4267
Db 541 CTCTGATGGAAGATATCTTGCCACCTACTTCAAACTGACAGCCACATTTCTTTTGGC 600

QY 4268 AGATGAACAGCTCACTGCTGGGAAGCATCGG-CATGCTGAATCGGCACCTCAGCTGGCG 4326
Db 601 AGATGAACAGCTCACTGCTGGGAAGCATCGGCCATGCTGAACCTCGGCACCTCAGCTGGCG 660

QY 4327 TGCAATTAACACCTACCAGGTGCCCC 4351
Db 661 TGCAATTAACACCTACCAGGTGCCCC 685
```

```
RESULT 8
BX507682      650 bp mRNA linear EST 04-SEP-2003
LOCUS      DKFZp686G03261_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION      DKFZp686G03261 5', mRNA sequence.
ACCESSION      BX507682
VERSION      BX507682.1 GI:32040573
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 650)
AUTHORS      Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
Wewes H.W., Well B., Amid C., Osanger A., Fobo G., Han M. and
Miemann S.
TITLE      EST (Poustka A., Albert R., Moosmayer P., Schupp I.,
Wellenreuther R., et al.)
JOURNAL      Unpublished (2003)
COMMENT      Contact: MIPS
```

```
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686G03261) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..650
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686G03261"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pRipEx2; Site_1: SfiI; Site_2: SfiIB;
cDNA-collection"
```

ORIGIN

```
Query Match      14.5%; Score 648.4; DB 5; Length 650;
Best Local Similarity 99.8%; Pred. No. 8.8e-176;
Matches 649; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3062 TGGAGGTGAGAGAACCGCACAGGCCCTGCTTCTGCGGAACCTGAGAAGATTGACAGG 3121
Db 1 TGGAGGTGAGAGAACCGCACAGGCCCTGCTTCTGCGGAACCTGAGAAGATTGACAGG 60

QY 3122 CAGGCAGAGAGAACCATTTGATGCTGGGCTCTCTTACTTACCTCAGTACATAGACCAG 3181
Db 61 CAGGCAGAGAGAACCATTTGATGCTGGGCTCTCTTACTTACCTCAGTACATAGACCAG 120

QY 3182 TCATATACCTCGAGTCACTCAGAAAGCCGCGAGACTATCACCAGGCTCTCTGATGCT 3241
Db 121 TCATATACCTCGAGTCACTCAGAAAGCCGCGAGACTATCACCAGGCTCTCTGATGCT 180

QY 3242 CAGGCGCTGAACAAAGTCCAGGAGGAGAGCATGACCTTGTGACGATGATCACCA 3301
Db 181 CAGGCGCTGAACAAAGTCCAGGAGGAGAGCATGACCTTGTGACGATGATCACCA 240

QY 3302 CTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAGAC 3361
Db 241 CTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAGAC 300

QY 3362 GGAAGCAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAATTTGTTGCTGAAA 3421
Db 301 GGAAGCAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAATTTGTTGCTGAAA 360

QY 3422 TTGAACCTCTTAAACTATTGACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGGATTG 3481
Db 361 TTGAACCTCTTAAACTATTGACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGGATTG 420

QY 3482 GTTGACTAGTGGTGGATCCAACTACTGCTGGCAGACATATTTGCAAGGCACTGACGT 3541
Db 421 GTTGACTAGTGGTGGATCCAACTACTGCTGGCAGACATATTTGCAAGGCACTGACGT 480

QY 3542 TTCTTCTGCTACAGCTCCAGGCCCAACTTCTCCACACAGCACTATCCGAAGAACAG 3601
Db 481 TTCTTCTGCTACAGCTCCAGGCCCAACTTCTCCACACAGCACTATCCGAAGAACAG 540

QY 3602 CCATTGATCTGATGGACGTGGGTTTCACTGTTTGGAGGCTTTACATGGATGTGTCCGCTG 3661
Db 541 CCATTGATCTGATGGACGTGGGTTTCACTGTTTGGAGGCTTTACATGGATGTGTCCGCTG 600

QY 3662 TTCTGATGGGGCTTCTCGAACTTTGTGCGGATGCGCGAGAAAACAACTTGGC 3711
Db 601 TTCTGATGGGGCTTCTCGAACTTTGTGCGGATGCGCGAGAAAACAACTTGGC 650
```

```
RESULT 9
CN314767
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1700532202315 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
CN314767
CN314767.1 GI:47331181
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 653)
Bradenberger, R., Wei, H., Zhang, S., Lei, S., Mudge, J., Fisk, G. J.,
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,
Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Bradenberger R
Regenerative Medicine
Geront Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbradenberger@geron.com
Insert length: 653 Std Error: 0.00.
Location/Qualifiers
1. .653
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, cell lines H1, H7, and
H9"
/clone_lib="GRN ES"
/notes="oligo dt primed, full-length enriched cDNA library
from undifferentiated hES cell lines H1 (p32), H7 (p29),
and H9 (p26) maintained in feeder-free conditions"

FEATURES
source
1. .653
Score 642; DB 7; Length 653;
Query Match 14.4%;
Best Local Similarity 99.8%; Pred. No. 6.4e-174;
Matches 653; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3449 CTCGAAGCTCTAGCCAAATTCCTGAGGGATTCGGGTGACTAGTGGTGGATCCAACTACT 3508
DB 1 CTCGAAGCTCTAGCC-AAATTCCTGAGGGATTCGGGTGACTAGTGGTGGATCCAACTACT 59
QY 3509 CGCTGGCCGACACATCTTGCAGGCACTGACGTTCTTCTGCTACAGCTCCAGCCCA 3568
DB 60 CGCTGGCCGACACATCTTGCAGGCACTGACGTTCTTCTGCTACAGCTCCAGCCCA 119
QY 3569 AACTCTCTCACACAGCACTATCCGAGACACCCATTGATCTGATGGCGTTCAGCTGCTCA 3628
DB 120 AACTCTCTCACACAGCACTATCCGAGACACCCATTGATCTGATGGCGTTCAGCTGCTCA 179
QY 3629 CTGTTTGGAGCGCTTACATGGATGTGTCGCTGTCTGATGGGCTTCTCGAACTTTGG 3688
DB 180 CTGTTTGGAGCGCTTACATGGATGTGTCGCTGTCTGATGGGCTTCTCGAACTTTGG 239
QY 3689 CCGATCCGAGAAACAACTTGGCCAAACATCAATGAGGTGGTTCCTTGGAGCCAGAGCTG 3748
DB 240 CCGATCCGAGAAACAACTTGGCCAAACATCAATGAGGTGGTTCCTTGGAGCCAGAGCTG 299
QY 3749 ACTCGGCGCTCTGGAGGCGATCGCTCTCGCTCATTCGCCCGCCAGACACCGGCT 3808
DB 300 ACTCGGCGCTCTGGAGGCGATCGCTCTCGCTCATTCGCCCGCCAGACACCGGCT 359
QY 3809 TCATCACCACCATAGCCAAAGAGGTACACAGATACGGCTCTTGGAGCAAAATACCCAAAT 3868
DB 360 TCATCACCACCATAGCCAAAGAGGTACACAGATACGGCTCTTGGAGCAAAATACCCAAAT 419
QY 3869 CACAGCAGAAATATGCACACAACTCTTGCACGAGCTAAAGGGGAAATTTTGGAGTCA 3928

Db 420 CACAGCAGAAATATGCACACAACTCTTGCACGAGCTAAAGGGGAAATTTTGGAGTCA 479
QY 3929 TTGAAATTTCTTATTTGAAAGATGCCACACAGATGTTCTGAGATCTTCTCGTGGAGTTATGG 3988
DB 480 TTGAAATTTCTTATTTGAAAGATGCCACACAGATGTTCTGAGATCTTCTCGTGGAGTTATGG 539
QY 3989 ACATCATTTATGTTACTGCTTGAAGGATCTTTAGTAAAGAAAGGCTTTCAGAGATGTT 4048
DB 540 ACATCATTTATGTTACTGCTTGAAGGATCTTTAGTAAAGAAAGGCTTTCAGAGATGTT 599
QY 4049 TCCAGCCCATCTGCAGTTCTACATGCTGACGCTATTATGAGCGGAATCACAGAA 4102
DB 600 TCCAGCCCATCTGCAGTTCTACATGCTGACGCTATTATGAGCGGAATCACAGAA 653

RESULT 10
BI855881
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

603383042F2 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5391706 5',
mRNA sequence.
BI855881
BI855881.1 GI:15996628
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 766)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1998 row: d column: 11
High quality sequence stop: 719.
Location/Qualifiers
1. .766
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5391706"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 87"
/notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

FEATURES
source
1. .766
Score 619.6; DB 4; Length 766;
Query Match 13.9%;
Best Local Similarity 94.9%; Pred. No. 2.2e-167;
Matches 729; Conservative 0; Mismatches 24; Indels 15; Gaps 8;

QY 3294 CATCACCCTGG-TTGCTTATCAAG-TGTCACCAAAATGAAAAATTTCTACATCTTAC 3351
DB 4 CATCACCCTGGATTGCTTATCAAGATGTCACCAAAATGAAAAATTTCTACATCTTAC 63
QY 3352 GAGGAA-AGACGGAAGCAAGCTACCGCTATTGTTTCTTGGAGTAATAGGAGCTGAAT 3410
DB 64 GAGGAAATAGCGAAGCAAGCTACCGCTATTGTTTCTTGGAGTAATAGGAGCTGAAT 123
QY 3411 TGGTGTGAANTGAACCTCTTAACATATTGACGAGACCTCGAAGCTCTAGCCAAATTC 3470
```

```
Db 124 GGTGTC-----TGAACCTCTAAACTATAGACGACGACCTCGAAGCTCTAGCCAAATACC 177
QY 3471 TGAGGGAATTCGGGTTGACTAGTGGTGGATCCAACTACTCTGCTGGCCAGACATACCTTGCAA 3530
Db 178 TGAGGGAATTCGGGTTGACTAGTGGTGGATCCAACTACTCTGCTGGCCAGACATACCTTGCAA 237
QY 3531 GGCACATGACGTTTCTCTGCTACAGGCTCCAGGCCCAAACTTCCTCCACACGACACTAT 3590
Db 238 GGCACATGACGTTTCTCTGCTACAGGCTCCAGGCCCAAACTTCCTCCACACGACACTAT 297
QY 3591 CCGAAGAACAGCCATTCGATTCGATTCGAGCGTGGGTTCACTGTTTGGAGCGCTTACATGGA 3650
Db 298 CCGAAGAACAGCCATTCGATTCGATTCGAGCGTGGGTTCACTGTTTGGAGCGCTTACATGGA 357
QY 3651 TGTGTCGCTGTTCTGATGGGGTTCCTCGAACTTTGTGCGGATGCGCGAGAAACAATTGC 3710
Db 358 TGTGTCGCTGTTCTGATGGGGTTCCTCGAACTTTGTGCGGATGCGCGAGAAACAATTGC 417
QY 3711 CAACATCAAAATGGGCTTCCTGAGCCGACGAGCTGACTCGGCCGCTCTGCGAGGCA 3770
Db 418 CAACATCAAAATGGGCTTCCTGAGCCGACGAGCTGACTCGGCCGCTCTGCGAGGCA 477
QY 3771 TGCCCTCTCGCTCATTTGCCACCGCCAGACCAACCGGCTTCATCACCATAGCCAAAGA 3830
Db 478 TGCCCTCTCGCTCATTTGCCACCGCCAGACCAACCGGCTTCATCACCATAGCCAAAGA 537
QY 3831 GGTACACAGACATACGGCTCTTCAGCAAAATACCCAAATCACAGCAGAAATGACACAAC 3890
Db 538 GGTACACAGACATACGGCTCTTCAGCAAAATACCCAAATCACAGCAGAAATGACACAAC 597
QY 3891 AACTCTTGCACGAGCTTAA--GGGAAATTTTCAGAGTCAATTCGAAATTTCTTATGGA-AAA 3947
Db 598 AACTCTTGCACGAGCTTAAAGGGGACAAATCTCAGAGTCAATTCGAAATTTCTTATGCAA 657
QY 3948 GATGCCCCACAGATGTTGT--GGATCTTCTCTGAGAGGTTATGACATCATTTATGACTGCC 4006
Db 658 GATGCCCCACAGATGTTGTGGGATCTTCTGTTGGAAGATGAGACATCATTTATGACTGCC 717
QY 4007 TTGAA--GGATCTTATGTTAAAGAAAGGCTTTCAGAAATGTTTCCC 4052
Db 718 TTGAAAGGATCTTAGTTAAACAGAAAGGCTTTCAGAAAGATGTTTCCC 765

RESULT 11
CA328908
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CA328908 768 bp mRNA linear EST 09-JUL-2003
UI-M-FY0-cdc-o-23-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
IMAGE:6827856 5', mRNA sequence.
CA328908
CA328908.1 GI:24547006
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 768)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: gcapbs-i@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
source
1..768
```

```
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6827856"
/tissue_type="whole brain"
/dev_stages="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site:1: EcoR I;
Site:2: Not I; The library was constructed according
Bonalito, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
```

ORIGIN

Query Match	13.7%;	Score	613.4;	DB 6;	Length	768;			
Best Local Similarity	87.4%;	Pred. No.	1.4e-165;						
Matches	671;	Conservative	0;	Mismatches	97;	Indels	0;	Gaps	0;

QY 1775 GTGCAATTCGATCGTTGTGTGATGGGATAACAGCAGTTCAGAGATTCCTAAACGCTTGTGTG 1834

Db 1 GTGCGCTGATGTTGTGCAATGGGAAATCACAGCAGTGGAGATCTGACGCTTGTGTGAGC 60

QY 1835 AAGCTGTTCTCTGCTGCTGTTGATTCATCTTAGTCATCCAGAGTCAACCTTAAACAAGCTA 1894

Db 61 AAGCGTTCCTCGCGCTGTAGACTCTCTTAGTCATCCAGCAGTCAACCTTGAACAAGCCA 120

QY 1895 TGACGAGAGTGTGCTGCTCTTAAAAATATGGCCCATCATAGCTTACAAACCCCTTG 1954

Db 121 TGACAAGACGGAGTCTCGCGCGCTTTAAAAATATGGCCCATCATAGCTTGAACAACCCCTTG 180

QY 1955 CAACCTAACTCTTGGCTCTGAGGCACTGCAAGGGAAATTTTACCTAAATATTTCTCAT 2014

Db 181 CAACGAACTTTTGGCTCTGAGGCTCTGCAAGGGTAATTTTACTTAATATTCCTCAT 240

QY 2015 ACTCCCTGATGTTCAAGCAATAAAGACAAACCTAACAGACCCGGACATACATGTTGCTAT 2074

Db 241 ACTCCCTGATGTTCAAGCAATAAAGACAAACCTAACAGACCCGGATATCCATGTGCTTT 300

QY 2075 TCTTTGATGTGAAGCGTTGATTTTCAACTCTGACTCAAGAGCGCTCTAGGCCGAATA 2134

Db 301 TCTTTGATGTGAAGCGTTGATTTTCAACTCTGACTCAAGAGCGCTCTAGGCCGAATA 360

QY 2135 CTGCTCTTATTTCCCGAGAGAAATTTGCAAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCT 2194

Db 361 CTGCACTTATTTTCCCGAGAGAAATCTGCAAGAGCATCTGGCAGTTTCAGACAAAGGGGGCT 420

QY 2195 CTTTCTTAACTGGAAACCGAGCAGGCTTCTCTTCCAAACAGTGAAGAAACGATCAAG 2254

Db 421 CTTTCTTAACTGGAAACCGAGCAGGCTTCTTCTTCCAGCAAGTGAAGAAACATATCAAG 480

QY 2255 AGNACATCAAGAACACCTCTTGTATGATGAGAGGAGATGAGGAGATTAATCAGGCGAGA 2314

Db 481 AGAACAATAAAGGAACACCTTCTGATTTAGAGGAGGACCGAGGAGGGGGGAGGAGGAC 540

QY 2315 GAAGGGAAGAAAGTGTATCTGATATCGGTCCAGCAAAATCAAGCAATTTGACCTTATTAG 2374

Db 541 GCAGAGGATAGTAGCTGACCTGAGTACCGGCCAGCAAGTCCAGCCACTGACCTACTAG 600

QY 2375 AATATAATTTAACTATGGACACTGCAAGAGCTGTTTATGCTGCTCCCTTACCGCTGGGTT 2434

Db 601 AATACAACCTTACTATGGATACCGCAAAATTTATTCATGCTGCTCTCCAGCGCTGGGCT 660

QY 2435 TGAATGAAGTACTGATGAAGTTTGGCTTGGATGCGCTTGGAAATGCTGAAACCCCACTGCA 2494
 DB 661 TGAATGAAGTTCTGGATGAAGTTTGTCTCGATGCGCTCGGAATGCTGAAACCGCACTGCA 720

QY 2495 CCGTATCGTTGGCTCTTGTGACAGAGGAGCCCATATGTCATGATGC 2542
 DB 721 CGGTGCTCTTGGTCTCTATCGAGAGAGGTCTATATGTCGCTGATGC 768

RESULT 12
 B0770731 780 bp mRNA linear EST 26-JUL-2002
 LOCUS UI-M-F10-byw-a-03-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
 DEFINITION IMAGE:6400010 5', mRNA sequence.
 B0770731
 ACCESSION B0770731.1 GI:21979205
 VERSION
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 780)
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 TITLE NIH-MGC <http://mgi.nci.nih.gov/>.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1..780
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6400010"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_F10"
 /notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCAGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
 Query Match 13.6%; Score 608.2; DB 5; Length 780;
 Best Local Similarity 88.1%; Pred. No. 4.4e-164;
 Matches 661; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 3724 GGGTGGCTTGAAGCCAGCAGTACTCGGCGGCTCTCGAGGAGTACCCCTCTGCTC 3783
 DB 1 GGGCTGGCTTGAAGCCCTGAGTACTCTGCGCGGCTCTCGCAAGGACGCGCTGTCTC 60

QY 3784 ATTGCCACCGCCAGACACCCCGCTTTCATCACCAACCATAGCCAAAGAGGTACACAGACAT 3843
 DB 61 ATGCCACACCGCCAGACACCCCGCTTTCATCACCAACCATAGCTAAGGAGGTCCACAGACAC 120

QY 3844 ACGGCTCTTGCAGCAATACCCCAATCACAGCAGATATGACACACCACTCTTGCAGCA 3903
 DB 121 ACGGCTCTTGCAGCAATACCCCAATCACAGCAGATATGACACACCACTCTTGCAGCA 180

QY 3904 GCTAAAGGGGAAATTTTGAGAGTCAITGAAATTTCTTATGAAAAGATGCCACACAGATGTT 3963
 DB 181 GCTAAAGGGGAAATTTTGAGAGTCAITGAAATTTCTTATGAAAAGATGCCACAGATGTT 240

QY 3964 GTGATCTTCTCTGAGGTTATGACATCATTTATGCTTCTGAGGATCTTTAGTT 4023
 DB 241 GTGATCTTCTCTGAGGTTATGACATCATTTATGCTTCTGAGGATCTTTAGTT 300

QY 4024 AAAAGAGAGGTTCTTCAAGATGTTTCCAGCCTCTGAGGTTCTTACATGCTCAGCTAT 4083
 DB 301 AAAAGAGAGGTTCTTCAAGATGTTTCCAGCCTCTGAGGTTCTTACATGCTCAGCTAT 360

QY 4084 TATGAGCGGAATCACAGAAATAGCAGTTGGAGCTCGGCATGTTTCTAGTGGCCCTGTACGAC 4143
 DB 361 TATGAGCGGAATCACAGAAATAGCAGTTGGAGCTCGGCATGTTTCTAGTGGCCCTGTACGAC 420

QY 4144 ATCCGACTGGAAATGTTCAGCAATTCATGGACACCAAGGACCAATCATCTGAGTGGCT 4203
 DB 421 ATCCGACTGGAAATGTTCAGCAATTCATGGACACCAAGGACCAATCATCTGAGTGGCT 480

QY 4204 TTGCTCTGTATGGAAGATATCTTGGACCTACTTCAACACTGACAGCCACATTTCTTTT 4263
 DB 481 TTGCTCTGTATGGAAGATATCTTGGACCTACTTCAACACTGACAGCCACATTTCTTTT 540

QY 4264 TGGCAGATGAACACGTCCTCTGCGAAGCATCGGCATGCTGAATCTCGGCACCTCAGCTG 4323
 DB 541 TGGCAGATGAACACGTCCTCTGCGAAGCATCGGCATGCTGAATCTCGGCACCTCAGCTG 600

QY 4324 CGCTGATTAATAACCTTACAGAGTGGCCCTCTGTGTCAGCCCGCTCCCGCTCCCAACAT 4383
 DB 601 CGCTGATTAATAACCTTACAGAGTGGCCCTCTGTGTCAGCCCGCTCCCGCTCCCAACAT 660

QY 4384 GGCCTCAAGTGGCCCGGCTCATCTGAGACTTCCAAACCGCAAGTCACTCTCATGGGCCAT 4443
 DB 661 GGCCTCAAGTGGCCCGGCTCATCTGAGACTTCCAAACCGCAAGTCACTCTCATGGGCCAT 720

QY 4444 GACGGGAAGGAGCAGCCGCTTCATGCTCTAA 4473
 DB 721 GATGGGAGGAGCAGCCGCTTCATGCTCTGA 750

RESULT 13
 B0770731 603 bp mRNA linear EST 01-MAR-2004
 LOCUS DKFp781L11961_r1 781 (synonym: hlec4) Homo sapiens cDNA clone
 DEFINITION DKFp781L11961 5', mRNA sequence.
 B0770731
 ACCESSION B0770731.1 GI:43437124
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 603)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 TITLE Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
 JOURNAL EST (Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., et al.)
 COMMENT Unpublished (2003)
 Contact: MIPS
 MIPS Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de

sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing Consortium of the German Genome Project.

No s1 sequence available.

This clone (DKFZp781l1961) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1. 603

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp781l1961"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="781 (synonym: hlcc4)"

/notes="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;

cDNA-collection"

ORIGIN

```
Query Match      13.4%; Score 601.4; DB 5; Length 603;
Best Local Similarity 99.8%; Pred. No. 3.7e-162;
Matches 602; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2385 AACTATGGACACATGCAAGCTGTTATGCTCTGCTTACGCTGGGGTTTGAATGAAGT 2444
DB 1 AACTATGGACACATGCAAGCTGTTATGCTCTGCTTACGCTGGGGTTTGAATGAAGT 60

QY 2445 ACTGGATGAAGTTGCTGGATCGCTTGGAAATGCTGAAACCCCATGCAACCGTATCGTT 2504
DB 61 ACTGGATGAAGTTGCTGGATCGCTTGGAAATGCTGAAACCCCATGCAACCGTATCGTT 120

QY 2505 TGGCTCTTGTCAAGAGGAGCCATATGTCATGATGCTGCGCGGTATTAATCAGCTGC 2564
DB 121 TGGCTCTTGTCAAGAGGAGCCATATGTCATGATGCTGCGCGGTATTAATCAGCTGC 180

QY 2565 TTGTAAACTCTCATATGGGAAACAGAAATGAGGAAGAGCTGCCAGCGTCTGAGGAGT 2624
DB 181 TTGTAAACTCTCATATGGGAAACAGAAATGAGGAAGAGCTGCCAGCGTCTGAGGAGT 240

QY 2625 AGAAAGGAACTTACGAGTGTCGCGTGCAGTCAACACAGATCTCTGCTATCAT 2684
DB 241 AGAAAGGAACTTACGAGTGTCGCGTGCAGTCAACACAGATCTCTGCTATCAT 300

QY 2685 TTCTTTGGCAAACTTTATATGATGATGACCAATGCACTTTTATGTTGATCATATGAA 2744
DB 301 TTCTTTGGCAAACTTTATATGATGATGACCAATGCACTTTTATGTTGATCATATGAA 360

QY 2745 GAAGGGTCTTACAGGCACTAGACCAAGCACCAGACCTTTCTAAGGCAAGGGGTTTC 2804
DB 361 GAAGGGTCTTACAGGCACTAGACCAAGCACCAGACCTTTCTAAGGCAAGGGGTTTC 420

QY 2805 CCCTCCAACTTCCAGTAATATTTGCAAGACAGATTAACAAGTTGCTGCACTGTGCT 2864
DB 421 CCCTCCAACTTCCAGTAATATTTGCAAGACAGATTAACAAGTTGCTGCACTGTGCT 480

QY 2865 TTCCGCTCGTGTGATGCTCATCTGCTGCTGACCTCTGACCTCTGCTCTGCTTTACA 2924
DB 481 TTCCGCTCGTGTGATGCTCATCTGCTGCTGACCTCTGACCTCTGCTCTGCTTTACA 540

QY 2925 TACTCTGTTTCTTAGTAATGAAGTTGGAGTCAAGTGTAGCTGTATGACATGTGTATGCT 2984
DB 541 TACTCTGTTTCTTAGTAATGAAGTTGGAGTCAAGTGTAGCTGTATGACATGTGTATGCT 600

QY 2985 GCC 2987
DB 601 GCC 603
```

RESULT 14
CR544624
LOCUS
DEFINITION DKFZp459G1824_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
631 bp mRNA linear EST 07-JUL-2004
DKFZp459G1824_5', mRNA sequence.

ACCESSION

CR544624

CE544624.1 GI:49896285

VERSION

EST.

KEYWORDS

Pongo pygmaeus (orangutan)

SOURCE

Pongo pygmaeus

ORGANISM

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Pongo.

REFERENCE

1 (bases 1 to 631)

AUTHORS

Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,

Fobo,G., Han,M. and Wiemann,S.

TITLE

Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)

JOURNAL

Unpublished (2004)

COMMENT

Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert from S. Wiemann,

Molecular Genome Analysis, German Cancer Research Center (DKFZ);

Email s.wiemann@dkfz-heidelberg.de; sequenced by BMF (Biomedical

Research Center at the Heinrich-Heine-University,

Duesseldorf/Germany) within the cDNA sequencing consortium of the

German Genome Project. This clone (DKFZp459G1824) is available at

the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,

Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:

clone@rzpd.de Further information about the clone and the

sequencing project is available at

http://mips.gsf.de/projects/cdna/.

FEATURES

Location/Qualifiers

1. 631

/organism="Pongo pygmaeus"

/mol_type="mRNA"

/db_xref="taxon:9600"

/clone="DKFZp459G1824"

/tissue_type="cortex"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="459 (synonym: pcor1)"

/note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB"

ORIGIN

Query Match

13.4%; Score 601.2; DB 7; Length 631;

Best Local Similarity 97.8%; Pred. No. 4.3e-162;

Matches 609; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3400 GGAGCTGAATTTGGTCTGAAATTTGAACCTCTTAACTATTGACACAGACTCGAAGCTCT 3459

DB 9 GGGCTGAATTTGGTCTGAAATTTGAACCTCTTAACTATTGACACAGACTCGAAGCTCT 68

QY 3460 AGCCAAATTCCTGAGGATTCGGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGA 3519

DB 69 AGCCAAATTCCTGAGGATTCGGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGA 128

QY 3520 CATACTGGACGACTGAGTTTCTTCTGCTACAGCTTCCAGCCCAAGCCCAAACTTCTCCA 3579

DB 129 CATACTGGACGACTGAGTTTCTTCTGCTGAGCCCAAGCCCAAACTTCTCCA 188

QY 3580 CACAGCACTATCCGAGAACAGCATTGATCTGATTGGACGTGGGTTCACTGTTGGGAG 3639

DB 189 CACAGCACTATCCGAGAACAGCATTGATCTGATTGGACGTGGGTTCACTGTTGGGAG 248

QY 3640 CCTTACATGGATGTGTCGCTGTTCTGATGGGCTTCTCGAACTTTGTGCCGATGCCGAG 3699

DB 249 CCTTACATGGATGTGTCGCTGTTCTGATGGGCTTCTCGAACTTTGTGCCGATGCCGAG 308

QY 3700 AAACAACTTCCCAACATCACAATGGGTTGCTCTGAGCCGACGACTGATCGGCCCGC 3759

DB 309 AAACAACTTCCCAACATCACAATGGGTTGCTCTGAGCCGACGACTGATCGGCCCGC 368

QY 3760 TCTGCGAGGACATGCCCTCTCGCTCATTTGCCCGCCAGACACCCCGCTTCAATCACCACC 3819

DB 369 TCTGCGAGGACATGCCCTCTCGCTCATTTGCCCGCCAGACACCCCGCTTCAATCACCACC 428

QY 3820 ATAGCCAAAGAGGTACACAGACATACGGCTCTTTGACGAGCAAAATCCCAATCAGACGAAT 3879

Db 429 ATAGCCAAAGAGGTACACAGACATACGGCTCTTGACGCAATACCCCAATACAGCAGAAAT 488

Qy 3880 ATGCACACAACTCTTGCACGAGTAAAGGGGAAATTTTGAGAGTCATTGAAATCTTT 3939

Db 489 ATGCACACAACTCTCGCACGCGCTAAAGGGGAAATTTTGAGAGTCATTGAAATCTTT 548

Qy 3940 ATTGAAAGATGCCACAGATGTTGGATCTTCTCGTGAGGTATTGACATCATTTATG 3999

Db 549 ATTGAAAGATGCCACGAGTGTGGATCTTCTTGTGGAGGTATTGACATCATTTATG 608

Qy 4000 TACTGCCTTGAAGGATCTTTAGT 4022

Db 609 TACTGCCTTGAAGGATCTTTAGT 631

RESULT 15

CF535467

LOCUS

DEFINITION UI-M-GHO-cho-g-17-0-UI.r1 NIH_BMAP_GHO Mus musculus cDNA clone

IMAGE:30362368 5', mRNA sequence.

CF535467

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF535467 714 bp mRNA linear EST 12-SEP-2003

UI-M-GHO-cho-g-17-0-UI.r1 NIH_BMAP_GHO Mus musculus cDNA clone

IMAGE:30362368 5', mRNA sequence.

CF535467

EST.

CF535467.1 GI:34587435

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 714)

NIH-MGC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-i@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/mousefl.html>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pyX-5.

Location/Qualifiers

FEATURES

source

1..714

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30362368"

/tissue_type="Whole brain"

/dev_stage="1, 5, and 15 days newborn"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_GHO"

/note="Organ: Brain; Vector: pyX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according

to the protocol of Benton and Murray (1989) and

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with NotI and then cloned

directionally into pyX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is CGAATCTGAT. This library was created for the University

of Iowa Brain Anatomy Project (BMAP): 'Gene discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,

program coordinator."

ORIGIN

Query Match 13.3%; Score 593.2; DB 7; Length 714;

Best Local Similarity 90.3%; Pred. No. 9.6e-160;

Matches 645; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

Qy 1433 CTCGGTATGATCAAGATACCTGATATCTGGAGGTGTGATTTTTCAGTCATAA-TTTGG 1491

Db 1 CTCGGTATGATCAAGATACCTGATATCTGGAGGTGTGATTTTTCAGTCATAA-TTTGG 60

Qy 1492 GACATATTTTCTGGAGAAATGAAACATATCTTCTGTGTTTCATGTTGGTGTGATTAATCTCAA 1551

Db 61 GACATATTTTCTGGAGAAATGAAACATATCTTCTGTGTTTCATGTTGGTGTGATTAATCTCAA 120

Qy 1552 CTTCTAGTTTCCACCTGAAACCTGTAGTCAAGAGTACAGCACTGTCATCTGCTCTGTAGCC 1611

Db 121 CTTCTGTTCCCGCCGGAACCTGTAGTCAAGAGTTCACACCTGTCATCTGTTCCGTAGCC 180

Qy 1612 AGTGACCACTCAGTAGGAATTTCTAAGTTTTCGAGAGAGAAAAATGCAATAATGTTGGCATCT 1671

Db 181 AGTGACCACTCAGTAGGAATTTCTAAGTTTTCGAGAGAGAAAAATGCAATAATGTTGGCATCT 240

Qy 1672 CGTCACTTTTTCCTATTCAAGTAAATGAGGCGCTTCTGATGATTAATACCTGTTGGTG 1731

Db 241 CGCCACCTCTTTTCTTATTCAAGTAAATGAGGCGCTTCTGATGATTAATACCTGTTGGTG 300

Qy 1732 GGGTGTTCAGATGGTGTCTGTGTACGTCTGCGAAATGGAATCTGGTGCATTTGGAATCTGTTGT 1791

Db 301 GGGTGTTCAGATGGTGTCTGTGTACGTCTGCGAAATGGAATCTGGTGCATTTGGAATCTGTTGT 360

Qy 1792 GTGATGGGATTAACAGCACTTGAAGTCTTAAACGCTTTGATGAAAGCTGTGTTCTGCTGCT 1851

Db 361 GCAATGGGAATCACAGCAGTGGAGATCTGAACGCTTTGTGCGAAGCCGTTCTGCGGCT 420

Qy 1852 GTTGATTCATTTAGTTCATCCAGAGTCAACCTTAAACAGCTATGACGAGACGTAGTCTT 1911

Db 421 GTAGACTCTCTTGTATCATCCAGCGTCAACCTGAAACAGCCATGACAGACGAGGTCTC 480

Qy 1912 GCTGCTCTTAAAAATATGCCCATCATAGCTACAAACCTTGCACCTAACTCTTTGGCT 1971

Db 481 GCGCCCTTAAATATATGSCCACCATAGCTGCAACCCCTTGCACAGAACCTTTTGGCT 540

Qy 1972 TCTGAGGCATCTGCAAGGGAATTTTACCTAAATATTTCTATAAATTTCTCATAAATCTCCCTGATGTTCAA 2031

Db 541 TCTGAGGCCTCTGCAAGGGAATTTTACCTAAATATTTCCATAAATTTCCATAAATCTCCCTGATGTTCAA 600

Qy 2032 GCAATTAAGACAAACCTTAACAGACCCGACATACATGCTGCTATTCTTTGATGTTGGAAGCG 2091

Db 601 GCAATTAAGACAAACCTTAACAGACCCGACATACATGCTGCTATTCTTTGATGTTGGAAGCT 660

Qy 2092 TTGATTTATTCAACTCTGACTCAAGAGGCTCTTAGSCCGAATACCTGCTCTTATT 2145

Db 661 TTGATTTATTCAACTCTGACTCAAGAGGCTCTTAGSCCGAATACCTGCTCTTATT 714

Search completed: July 2, 2005, 12:49:48

Job time : 12560 secs

THIS PAGE BLANK (USP 10)